Spectral Flow Cytometry:

A Comprehensive Tool for Deep Profiling of Cell Populations

Article Collection

CURRENT PROTOCOLS
A Wiley Brand

Cytometry PART A
ISAC
Journal of Quantitative Cell Science

Sponsored by
SONY

SONY

Focus on Discovery Research

With Reliable Operation and Highest Data Quality

Explore the advantages of innovative spectral cell analysis from Sony Biotechnology that supports a wide variety of applications, designed for ease of adoption and reliable operation—which are critical for today's multi-user environments.

ID7000™ Spectral Cell Analyzer

- **Advanced design for your high parameter experiments -** Configurable with up to 7 lasers, 186 detectors.
- **Ready for future fluorochrome development -** The ability to expand your multi-color panels, so you can lead today and be prepared for tomorrow.
- **Industry-leading AutoSampler and intuitive software workflows -** Enhance your laboratory's operations, efficiency, and productivity.

Explore the features of the ID7000 that will help you stay ahead today and excel tomorrow.

go.sonybiotechnology.com/core-labs

Contents

Imprint

©John Wiley & Sons, Inc.
111 River Street,
Hoboken, NJ 07030-5774
USA
Contact: Customer Service

Editor:
Róisín Murtagh

Senior Account Manager:
Joseph Tomaszewski

Sony Biotechnology
https://www.sonybiotechnology.com/

Cover: ©2024 Sony Biotechnology

Introduction

Flow cytometry is a technology used to measure the physical and chemical features of individual cells or particles and provides information about the size and complexity of the cells. In brief, a sample of fluid-suspended particles is injected into the instrument and passed, single-file, through a laser beam. As the cells pass through the laser, they absorb, refract, and emit light which is detected by an array of detectors. Furthermore, flow cytometry allows the identification and quantification of an abundance of specific cell types by attaching fluorophores to monoclonal antibodies that recognize cell-type specific proteins. Spectral flow cytometry goes a step further as it provides a "spectral fingerprint", allowing for spectral unmixing and the detection of many more fluorophores simultaneously via multiple lasers, with each fluorophore being detected by multiple detectors. This allows for the detection of over 35 markers from a single tube assay, improving the resolution and increasing the ability to distinguish hard-to-identify cell populations within the sample. In clinical laboratories, the technology is most used to support the diagnosis of immunodeficiencies, hematologic malignancies, and responses to therapy; to monitor the progress of stem cell transplants; and to manufacture CAR T cells.

The first commercial spectral flow cytometer was developed by Sony. The device originally used an array of prisms to disperse the collected light across the sensors. However, the newer model, the ID7000 analyzer, now uses a grating for a more uniform dispersion of light across the sensor arrays. This offers several advantages over conventional flow cytometry, including the ability to detect more fluorophores, the ability to generate deeper phenotypic datasets through stitched panels, and increased accuracy from eliminating autofluorescence as a variable. Additionally, spectral flow cytometry has the potential to enable systems biology analyses of immune cell networks that could reveal hidden patterns or features of immune system dysregulation. Spectral flow cytometry also has shorter set-up times and is easier to use in research settings.

This collection of peer-reviewed articles aims to showcase how spectral flow cytometry is a powerful and versatile technology for analyzing complex cell populations. The collection begins with a study by Nolan [1] that reviews the origins, current state, and future of spectral flow cytometry. He discusses how the technology has evolved from being used by individual researchers to being adopted by the mainstream as a way of analyzing the multiparameter immunophenotyping of immune cells. He also highlights the current instrumentation and software available and speculates on how the technology could be used in the future.

In a review article by Novo [2], the differences in data acquisition between traditional and spectral flow cytometers are explained, and how these differences allow for more accurate data analysis through the use of mathematical tools referred to as 'unmixing'. The article also provides an overview of the different mathematics and theories between traditional compensation and unmixing to better explain how the use of newer methods can provide a more in-depth analysis of data.

Finally, Monard [3] reports on the various challenges associated with using fluorescent proteins (FPs) in biological research and the potential solutions to these challenges. He explains how the introduction of full spectrum flow cytometers has allowed for the separation of at least six FPs, but the laser wavelengths of commercial instruments are not ideal for all FPs. He discusses the need for single-color controls and the expense and inconvenience associated with producing colonies of animals expressing each FP. Finally, he describes a procedure that can be used to produce and purify FPs and couple them to polystyrene microspheres, which can be stored and used without any special equipment or skills.

Through the concepts and applications presented in this article collection, we hope to educate scientists on the recent advances in the use of this technology. For more information, we encourage you to visit Sony to explore more options to enhance your research.

Róisín Murtagh
Editor at *Wiley Analytical Science*

References

[1] Nolan, J.P. (2022). The evolution of spectral flow cytometry. *Cytometry Part A.* DOI: 10.1002/cyto.a.24566.

[2] Novo, D. (2022). A comparison of spectral unmixing to conventional compensation for the calculation of fluorochrome abundances from flow cytometric data. *Cytometry Part A.* DOI: 10.1002/cyto.a.24669.

[3] Monard, S. (2022). Building a spectral cytometry toolbox: Coupling fluorescent proteins and antibodies to microspheres. *Cytometry Part A.* DOI: 10.1002/cyto.a.24557.

Received: 6 December 2021 | Revised: 13 April 2022 | Accepted: 22 April 2022

DOI: 10.1002/cyto.a.24566

REVIEW ARTICLE

ISAC CYTOMETRY Journal of Quantitative Cell Science PART A

The evolution of spectral flow cytometry

John P. Nolan

Scintillon Institute, San Diego, California, USA

Correspondence
John P. Nolan, Scintillon Institute, 6868 Nancy Ridge Dr, San Diego, CA 92121, USA.
Email: jnolan@scintillon.org

Abstract

This special issue of *Cytometry* marks the transition of spectral flow cytometry from an emerging technology into a transformative force that will shape the fields of cytometry and single-cell analysis for some time to come. Tracing its roots to the earliest years of flow cytometry, spectral flow cytometry has evolved from the domain of individual researchers pushing the limits of hardware, reagents, and software to the mainstream, where it is being harnessed and adapted to meet the analytical challenges presented by modern biomedical research. In particular, the current form of spectral flow technology has arisen to address the needs of multiparameter immunophenotyping of immune cells in basic and translational research, and much of the current instrumentation and software reflects the needs of those applications. Yet, the possibilities enabled by high-resolution analysis of the spectral properties of optical absorbance, scatter, and emission have only begun to be exploited. In this brief review, the author highlights the origins and early milestones of single-cell spectral analysis, assesses the current state of instrumentation and software, and speculates as to future directions of spectral flow cytometry technology and applications.

KEYWORDS
detector array, dispersive optics, excitation-emission matrix, unmixing

1 | INTRODUCTION

This special issue of *Cytometry* marks the transition of spectral flow cytometry from an emerging technology into a transformative force that will shape the fields of cytometry and single-cell analysis for some time to come. Tracing its roots to the earliest years of flow cytometry [1], spectral flow cytometry has evolved from the domain of individual researchers pushing the limits of hardware, reagents, and software to the mainstream, where it is being harnessed and adapted to meet the analytical challenges presented by modern biomedical research. In particular, the current form of spectral flow technology (referred to as "Full Spectrum" Flow Cytometry in the title of this Issue) has arisen to address the needs of multiparameter immunophenotyping of immune cells in basic and translational research, and much of the current instrumentation and software reflects the needs of those applications. Yet, the possibilities enabled by high-resolution analysis of the spectral properties of optical absorbance, scatter, and emission have only begun to be exploited, and the field is still in its infancy. In this brief review, the author highlights the origins and early milestones of single-cell spectral analysis in flow, assess the current state of instrumentation and software, and speculate as to future directions of spectral flow cytometry technology and applications.

2 | BEGINNINGS

The roots of spectral flow cytometry, as for many things fluorescence, can be traced to the work of Gregorio Weber [2], who pioneered the use of fluorescence to study biological systems. Weber, working at the Universities of Sheffield and Illinois, established the conceptual and practical underpinnings of bioanalytical fluorescence spectroscopy, integrating chemistry, physics, and engineering to produce dyes, instruments, and experimental approaches to study systems ranging from proteins to cells [3]. At about the time Fulwyler was building the first single-cell sorter [4], and some years before the first fluorescence flow cytometers [5–7], Weber, recognizing the wealth of information

contained in the excitation and emission spectra of fluorescent compounds, described the determination of the number and abundances of multiple fluorescent species in a mixed sample using the excitation-fluorescence (EF) matrix [8]. In the years that followed, this approach, most often referred to as excitation-emission matrix spectroscopy, inspired the development of new instrumentation for rapid spectral analysis [9, 10] and data analysis treatments [11, 12] that were foundational for the fields of chemometrics and biomedical spectroscopy. More than 50 years later, these principles can now be applied to measurements of individual cells.

3 | EARLY WORK

The first published attempts to measure the spectra of individual cells in flow was reported by Wade and colleagues [13] who used grating-based spectrometers to disperse light collected from a flow cell onto vidicon detectors, arrays of silicon-intensified photon counters that provided sensitive detection with (relatively) fast single integration times (~10 ms). These systems were used to measure the chlorophyl autofluorescence of blue green algae and cultured mammalian cells stained with the nucleic acid stain propidium iodide and fluorescamine, a fluorogenic amine-reactive dye. The systems were operated in a "continuous" mode, in which the signal from many cells was accumulated to obtain a total, or average, spectra of all the cells that passed the detector (~20,000 cells in ~20 s), and in "gated" mode, in which data acquisition was triggered by the signal from a separate photomultiplier detector to measure the spectra of individual cells. These demonstrations served as proofs of principle but were ultimately limited by the cycle speed of the vidicon spectral detectors available at the time.

In the years that followed, there were several attempts to adapt fluorescence spectral detection capabilities of flow cytometer-based instruments. Steen and Stokke [14, 15] used a scanning monochromometer adapted to a commercial flow cytometer to make sequential measurements of cells at different emission wavelengths to produce an average spectrum of a cell population stained with Hoechst 33258. A decade later, Asbury and van den Engh [16] reported on a similar, monochromometer-based approach to measure the spectra of sperm cells stained with a number of different nucleic acid stains. At Los Alamos, Buican [17] developed a Fourier-transform flow cytometer that used a high-speed interferometric approach to determine the intensity spectra of individual particles in a flow system, though in practice limited signal required signal averaging of many particles. Both of these approaches employed PMTs as detectors, which enabled the rapid measurement of many cells, but interrogating only one emission band at a time.

By the mid-1990s, steps toward the precursors of the modern spectral flow cytometers were in evidence. Gauci and colleagues [18] used a prism to disperse light for the flow cell over a 512-element intensified photodiode array, triggered by a light scatter signal, to measure the spectra of alignment beads, as well as individual Dictyostelium cells stained with FITC, PE, or Cy3. The rate of measurement of individual particles in this system was limited by the frame rate (62.5 Hz) of the detector.

Fuller and Sweedler [19] used a grating to disperse light over a CCD array to detect the spectra of individual synthetic lipid vesicles prepared with fluorescein- or rhodamine-labeled lipids, excited by two different excitation lasers. The CCD format, 1024 × 256 pixels, provided sub-nanometer spectral resolution and was operated in a continuous mode, and the detector output was analyzed post-acquisition to identify events, which were identified based on their emission spectra.

Dubelaar [20] integrated a grating spectrometer with a multipixel hybrid PMT array to measure light scatter and fluorescence from algae in an autonomous flow cytometer designed for remote measurement of phytoplankton in seawater. This detector employed a seven-pixel array but used only three of the pixels to measure light scatter and two fluorescence emission bands. This instrument also recorded the signal pulse shape of each event, for each channel.

While these flow cytometry instrument developments were occurring, flow cytometry applications, especially multicolor immunophenotyping of lymphocytes and other immune cells [21, 22] were driving the development of conventional flow cytometer instruments to ever higher number of lasers and detectors. This, along with an expansion in the number of different fluorescent conjugates for antibody labeling and software tools to facilitate data analysis [23, 24], characterized so-called polychromatic flow cytometry [25]. Limitations in the numbers of probes that could be resolved by conventional flow cytometers inspired mass cytometry using lanthanide-conjugated antibodies [26] and, ultimately, the spectral flow cytometers we see today.

4 | THE MODERN PERIOD

By the early 2000s, advances in detector technology began to provide both the speed and resolution required for practical application in cytometry applications. PMTs were the photodetector of choice for demanding applications like flow cytometry, owing to their high gain and fast response times. Robinson and colleagues [27–30] used a grating to disperse light over a 32-channel multianode PMT array to demonstrate high-speed single-cell spectral analysis. They used principal components analysis (PCA) and, later, least squares unmixing [30], to resolve differently stained particles and cells. The multianode PMT detector-based approach was later adapted by Sony in the first commercial spectral flow cytometer [31], which used an array of prisms to disperse collected light across the multianode PMT.

Multianode PMTs provide the characteristic advantages of PMT-based detection, including speed and high gain, but also limited quantum efficiency and a limited number of channels [32] making them less suitable for applications requiring high sensitivity or high spectral resolution. Spectroscopy-grade CCD-type detectors detectors, in contrast, generally provide higher quantum efficiency and a greater number (thousands) of detector elements in a high-density physical arrangement that enables high spectral resolution. Goddard and

colleagues used a volume phase holographic grating to disperse light over a 128 × 1024 pixel CCD array to demonstrate high QE (>80%), high-resolution (∼1 nm) spectral measurements of calibrated beads and propidium iodide-stained mammalian cells [32]. Subsequent refinement of this approach enabled measurement of high sensitivity and high-resolution (<1 nm) fluorescence and Raman spectra [33–37], including for measurement of SERS from individual Au and Ag nanoparticles [38] and the multicolor immunophenotyping of PBMCs using least squares unmixing [39]. These CCD-based systems could provide very high sensitivity and spectral resolution, though the readout speed of detectors available at the time generally limited particle measurement rates to ∼1000/s. Newer CCD (Andor iXon) and CMOS-based (Hamamatsu linear CMOS) detectors can support acquisition rates >10,000/s.

In the decade following release of the first commercial flow cytometer designed to enable spectral analysis, several additional instruments have come onto the market, each with a distinct hardware approach. The original Sony Spectral analyzer used an array of prisms to disperse the collected light across an array of regularly spaced sensors on a multianode PMT. Several new instruments have taken approaches that resemble conventional instruments in that they use dichroic filters to select emission bands that are detected using avalanche photodiodes (APDs), as in the Cytek Aurora and Northern Lights, or PMTs, as in the BD Symphony A5 SE and Thermo Big Foot. The latter instrument is a sorter, as is the Aurora CS, which demonstrates that the problem of performing unmixing in real-time to make sort decisions has been solved, at least for simple unmixing approaches. Meanwhile, the newest Sony ID7000 analyzer uses a grating, rather than prisms, for more uniform dispersion of collected emission across its sensor arrays.

The diversity of approaches illustrated by this current generation of spectral instruments highlights the reality that spectral flow cytometry is more about the data processing and analysis than the hardware used to collect the data. While the faithful representation of emission spectra provided by gratings and linear detector arrays is attractive and useful for spectroscopy-focused applications, simple unmixing to estimate the abundances of known spectral components in a mixture does not require that spectral resolution be high or uniform across the spectral range, and modest and variable spectral resolution is suitable for performing lymphocyte immunophenotyping, for example [40–45]. In fact, data from conventional instruments designed for use with traditional compensation can be analyzed in a "pseudo-spectral" manner [46], in which the signals from all detectors are used to form a spectrum (albeit of low and variable resolution) for each fluorochrome that can be used to unmix and determine the abundance of each fluorophore from a mixture spectrum. As Novo describes elsewhere in this issue (ref), compensation is a "square matrix" variant of more general spectral unmixing problems where there are more detectors than fluorophores/components, and unmixing can be applied to data from any instrument where this is the case. Conversely, some instruments designed for spectral analysis provide for the data to be saved in a "virtual filter" mode, where several individual spectral channels are combined to form a single intensity value that approximates that which would be obtained from a conventional instrument. In its more general form, unmixing of single particle emission spectra excited at multiple excitation wavelengths can be recognized as the single particle implementation of the excitation-emission matrix spectroscopy approach to fluorescence first described by Weber.

5 | THE FUTURE

Progress in the development and translation of any technology is driven by the needs of the market. While the roots of any transformative technology can be traced to the curiosity and interests of individual researchers, its further development into practical (and commercial) reality depends on its ability to solve important problems faced by significant numbers of users. High-dimensional cell analysis has long been a dominant driver for flow cytometry technology development, and its influence on spectral flow cytometry development is no exception. Now, as has been the case for much of the field's existence, most commercial flow cytometers are designed to immunophenotype lymphocytes using fluorescent antibodies and probes, and the current generation of spectral flow cytometers appear to excel at this. Much of the recently published work using spectral flow cytometry has focused on the optimization of immunophenotyping staining panels and protocols, and their validation by comparison with conventional polychromatic and mass cytometric approaches [40–47]. The advantages of simpler workflows and improved resolution compared to non-spectral analysis are spurring rapid adoption in academics and industry [47], and we might expect future spectral flow cytometry development to address outstanding challenges in multicolor immunofluorescence not possible with a conventional "square-matrix" approach.

Among the challenges that arises in high-dimensional flow cytometry is the deviation of a particular conjugate from its ideal or typical spectrum. For example, tandem conjugates can decompose [48] such that their spectra change, and unexpected probe-probe interactions between molecules bound in or on a cell can confound linear unmixing models that assume static component spectra. However, if the deviation from ideal can be measured and understood, it should be possible to apply fitting algorithms that account for this behavior using alternating least squares or other approaches that allow the base spectra to vary, within constraints [49]. Such approaches might be the basis of algorithms that could accommodate some of the common sources of immunofluorescent conjugate variation.

Cellular autofluorescence has long been viewed as an undesired source of background that interferes with the signal from dim fluorophores and/or low abundance markers, and much effort has been directed at "correcting" measurements to account for autofluorescence [50–54]. Another perspective considers that cellular autofluorescence, which can arise from several endogenous metabolites, amino acids and other molecules [55–57], is a rich source of information about cell state [58, 59]. Spectral measurement presents the opportunity to unify these perspectives by enabling the

estimation of the abundance of both endogenous intrinsic fluorophores and exogenous fluorescent conjugates. Work toward this is at a very early stage [41, 60], but unmixing approaches that consider individual spectra of major autofluorescence components would in principle enable those immunofluorescence fluorophores whose spectra overlapped to be detected at lower abundances.

Yet, immunophenotyping is only one cytometric measurement, and flow cytometry technology is useful for more than lymphocyte analysis. Fluorescence resonance energy transfer (FRET) can be used to estimate the proximity of fluorophores and/or fluorescent antibodies on or in a cell [61–63], has also been exploited to design intracellular molecular sensors whose emission spectra change upon analyte sensing [64–66]. Like immunofluorescence, FRET measurements can be compromised by autofluorescence [53], and spectral unmixing approaches may enhance high-resolution FRET measurements in the presence of other spectrally overlapping fluorescence signals [65].

Among the application areas that might be expected to drive the continued evolution of spectral flow cytometry are the resolution of dim signals from various sources of background. For quantitative measurements, sensitivity is generally limited by background and, for cells and other biological particles, the predominant background is intrinsic autofluorescence of various origins. This has implications for the measurement of low abundance, "dim" antigens on cells, but also for the detection of very low abundance targets (e.g., single molecule) on biological nanoparticles such as viruses, virus-like-particles (VLPs), and extracellular vesicles (EVs) [67]. For very dim particles, autofluorescence might be on the same order as optical and electronic noise, which may have their own spectral characteristics, and thus can be accounted for as either fixed or variable background components in an unmixing process. Moreover, signals from fluorophores of interest and from various sources of background can have their own distinctive variances, for example Gaussian-type noise distributions in sources of electronic background versus Poisson-dominated variance in dim signals from small number of photons produced by small numbers of labels. The accurate measurement of these background signals, and their variances, should improve fluorescence detection limits [30].

In conclusion, we can anticipate that future generations of flow cytometers, whether designed for very high-dimensional analysis of cells or for single molecule sensitivity and resolution, will be spectral instruments that operate on the full excitation-emission matrix that Weber described more than 50 years ago [8].

AUTHOR CONTRIBUTIONS

John P. Nolan: Writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

The author is a professor at the Scintillon Institute, where the research was supported by grants from the National Institutes of Health, and CEO at Cellarcus Biosciences, which provides flow cytometry-related products and services. He is an inventor on patents and patent applications related to flow cytometry.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/cyto.a.24566.

ORCID

John P. Nolan https://orcid.org/0000-0001-5845-3764

REFERENCES

1. Nolan JP, Condello D. Spectral flow cytometry. Curr Protoc Cytom. 2013;63:1.27.1–1.27.13.
2. Jameson DM. A fluorescent lifetime: reminiscing about Gregorio Weber. In: Jameson DM, editor. Perspectives on fluorescence: a tribute to Gregorio Weber. Cham: Springer International Publishing; 2016. p. 1–16.
3. Jameson DM. The seminal contributions of Gregorio Weber to modern fluorescence spectroscopy. In: Valeur B, Brochon J-C, editors. New trends in fluorescence spectroscopy: applications to chemical and life sciences. Berlin, Heidelberg: Springer; 2001. p. 35–58.
4. Fulwyler MJ. Electronic separation of biological cells by volume. Science. 1965;150:910–1.
5. Dilla MAV, Truiullo TT, Mullaney PF, Coultex JR. Cell Microfluorometry: a method for rapid fluorescence measurement. Science. 1969;163:1213–4.
6. Dittrich W, Göhde W. Impulsfluorometrie bei einzelzellen in suspensionen. Zeitschrift für Naturforschung B. 1969;24:360–1.
7. Hulett HR, Bonner WA, Barrett J, Herzenberg LA. Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence. Science. 1969;166:747–9.
8. Weber G. Enumeration of components in complex systems by fluorescence spectrophotometry. Nature. 1961;190:27–9.
9. Johnson DW, Callis JB, Christian GD. Rapid scanning fluorescence spectroscopy. Anal Chem. 1977;49:747A–57A.
10. Warner IM, Callis JB, Davidson ER, Gouterman M, Christian GD. Fluorescence analysis: a new approach. Anal Lett. 1975;8:665–81.
11. Neal SL, Davidson ER, Warner IM. Resolution of severely overlapped spectra from matrix-formatted spectral data using constrained nonlinear optimization. Anal Chem. 1990;62:658–64.
12. Leggett DJ. Numerical analysis of multicomponent spectra. Anal Chem. 1977;49:276–81.
13. Wade CG, Rhyne RH, Woodruff WH, Bloch DP, Bartholomew JC. Spectra of cells in flow cytometry using a vidicon detector. J Histochem Cytochem. 1979;27:1049–52.
14. Steen HB, Stokke T. Fluorescence spectra of cells stained with a DNA-specific dye, measured by flow cytometry. Cytometry. 1986;7:104–6.
15. Steen HB, Stokke T. Fluorescence spectra of Hoechst 33258 bound to chromatin. Biochim Biophys Acta Gene Struct Express. 1986;868:17–23.
16. Asbury CL, Esposito R, Farmer C, van den Engh G. Fluorescence spectra of DNA dyes measured in a flow cytometer. Cytometry. 1996;24:234–42.
17. Buican T. Real-time Fourier transform spectrometry for fluorescence imaging and flow cytometry. SPIE. 1990;1205:126–133.
18. Gauci MR, Vesey G, Narai J, Veal D, Williams KL, Piper JA. Observation of single-cell fluorescence spectra in laser flow cytometry. Cytometry. 1996;25:388–93.
19. Fuller RR, Sweedler JV. Characterizing submicron vesicles with wavelength-resolved fluorescence in flow cytometry. Cytometry. 1996;25:144–55.
20. Dubelaar GBJ, Gerritzen PL, Beeker AER, Jonker RR, Tangen K. Design and first results of CytoBuoy: a wireless flow cytometer for in situ analysis of marine and fresh waters. Cytometry. 1999;37:247–54.

21. De Rosa SC, Herzenberg LA, Roederer M. 11-color, 13-parameter flow cytometry: identification of human naive T cells by phenotype, function, and T-cell receptor diversity. Nat Med. 2001;7:245–8.
22. Perfetto SP, Chattopadhyay PK, Roederer M. Seventeen-colour flow cytometry: unravelling the immune system. Nat Rev Immunol. 2004; 4:648–55.
23. Roederer M. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. Cytometry. 2001;45:194–205.
24. Roederer M. Compensation in flow cytometry. Curr Protoc Cytom. 2002;22:1.14.1–1.14.20.
25. Chattopadhyay PK, Hogerkorp CM, Roederer M. A chromatic explosion: the development and future of multiparameter flow cytometry. Immunology. 2008;125:441–9.
26. Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem. 2009;81:6813–22.
27. Grégori G, Patsekin V, Rajwa B, Jones J, Ragheb K, Holdman C, et al. Hyperspectral cytometry at the single-cell level using a 32-channel photodetector. Cytometry A. 2012;81:35–44.
28. Robinson JP, Rajwa B, Gregori G, Jones J, Patsekin V. Multispectral cytometry of single bio-particles using a 32-channel detector. Proc. SPIE 5692, Advanced Biomedical and Clinical Diagnostic Systems III, 2005. p 359–365.
29. Grégori G, Rajwa B, Patsekin V, Jones J, Furuki M, Yamamoto M, et al. Hyperspectral cytometry. In: Fienberg HG, Nolan GP, editors. High-dimensional single cell analysis. Current Topics in Microbiology and Immunology. Volume 377. Berlin Heidelberg: Springer; 2014. p. 191–210.
30. Novo D, Grégori G, Rajwa B. Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices. Cytometry A. 2013;83A:508–20.
31. Futamura K, Sekino M, Hata A, Ikebuchi R, Nakanishi Y, Egawa G, et al. Novel full-spectral flow cytometry with multiple spectrally-adjacent fluorescent proteins and fluorochromes and visualization of in vivo cellular movement. Cytometry A. 2015;87:830–42.
32. Goddard G, Martin JC, Naivar M, Goodwin PM, Graves SW, Habbersett R, et al. Single particle high resolution spectral analysis flow cytometry. Cytometry A. 2006;69:842–51.
33. Watson DA, Brown LO, Gaskill DF, Naivar M, Graves SW, Doorn SK, et al. A flow cytometer for the measurement of Raman spectra. Cytometry A. 2008;73:119–28.
34. Watson DA, Gaskill DF, Brown LO, Doorn SK, Nolan JP. Spectral measurements of large particles by flow cytometry. Cytometry A. 2009;75:460–4.
35. Nolan JP, Duggan E, Condello D. Optimization of SERS tag intensity, binding footprint, and emittance. Bioconjug Chem. 2014;25:1233–42.
36. Nolan JP, Duggan E, Liu E, Condello D, Dave I, Stoner SA. Single cell analysis using surface enhanced Raman scattering (SERS) tags. Methods. 2012;57:272–9.
37. Nolan JP, Sebba DS. Surface-enhanced Raman scattering (SERS) cytometry. In: Paul M, editor. Methods in cell biology. Volume 102. San Diego, CA: Academic Press; 2011. p. 515–32.
38. Sebba DS, Watson DA, Nolan JP. High throughput single nanoparticle spectroscopy. ACS Nano. 2009;3:1477–84.
39. Nolan JP, Condello D, Duggan E, Naivar M, Novo D. Visible and near infrared fluorescence spectral flow cytometry. Cytometry A. 2012;83: 253–64.
40. Ferrer-Font L, Mayer JU, Old S, Hermans IF, Irish J, Price KM. High-dimensional data analysis algorithms yield comparable results for mass cytometry and spectral flow cytometry data. Cytometry A. 2020;97:824–31.
41. Ferrer-Font L, Pellefigues C, Mayer JU, Small SJ, Jaimes MC, Price KM. Panel design and optimization for high-dimensional Immunophenotyping assays using spectral flow cytometry. Curr Protoc Cytom. 2020;92:e70.
42. Fox A, Dutt TS, Karger B, Obregón-Henao A, Anderson GB, Henao-Tamayo M. Acquisition of high-quality spectral flow cytometry data. Curr Protoc Cytom. 2020;93:e74.
43. Park LM, Lannigan J, Jaimes MC. OMIP-069: forty-color full Spectrum flow cytometry panel for deep Immunophenotyping of major cell subsets in human peripheral blood. Cytometry A. 2020;97: 1044–51.
44. Schmutz S, Valente M, Cumano A, Novault S. Spectral cytometry has unique properties allowing multicolor analysis of cell suspensions isolated from solid tissues. PLOS One. 2016;11:e0159961.
45. Solomon M, DeLay M, Reynaud D. Phenotypic analysis of the mouse hematopoietic hierarchy using spectral cytometry: from stem cell subsets to early progenitor compartments. Cytometry A. 2020;97: 1057–65.
46. Jimenez-Carretero D, Ligos JM, Martínez-López M, Sancho D, Montoya MC. Flow cytometry data preparation guidelines for improved automated phenotypic analysis. J Immunol. 2018;200: 3319–31.
47. McCausland M, Lin Y-D, Nevers T, Groves C, Decman V. With great power comes great responsibility: high-dimensional spectral flow cytometry to support clinical trials. Bioanalysis. 2021;13:1597–616.
48. Hulspas R, Dombkowski D, Preffer F, Douglas D, Kildew-Shah B, Gilbert J. Flow cytometry and the stability of phycoerythrin-tandem dye conjugates. Cytometry A. 2009;75:966–72.
49. Heylen R, Zare A, Gader P, Scheunders P. Hyperspectral Unmixing with endmember variability via alternating angle minimization. IEEE Trans Geosci Remote Sens. 2016;54:4983–93.
50. Corsetti JP, Sotirchos SV, Cox C, Cowles JW, Leary JF, Blumburg N. Correction of cellular autofluorescence in flow cytometry by mathematical modeling of cellular fluorescence. Cytometry. 1988;9: 539–47.
51. Hulspas R, O'Gorman MR, Wood BL, Gratama JW, Sutherland DR. Considerations for the control of background fluorescence in clinical flow cytometry. Cytometry B Clin Cytometry. 2009;76:355–64.
52. Roederer M, Murphy RF. Cell-by-cell autofluorescence correction for low signal-to-noise systems: application to epidermal growth factor endocytosis by 3T3 fibroblasts. Cytometry. 1986;7:558–65.
53. Sebestyén Z, Nagy P, Horváth G, Vámosi G, Debets R, Gratama JW, et al. Long wavelength fluorophores and cell-by-cell correction for autofluorescence significantly improves the accuracy of flow cytometric energy transfer measurements on a dual-laser benchtop flow cytometer. Cytometry. 2002;48:124–35.
54. Steinkamp JA, Stewart CC. Dual-laser, differential fluorescence correction method for reducing cellular background autofluorescence. Cytometry. 1986;7:566–74.
55. Weber G. Fluorescence of riboflavin and flavin-adenine dinucleotide. Biochem J. 1950;47:114–21.
56. Teale F, Weber G. Ultraviolet fluorescence of the aromatic amino acids. Biochem J. 1957;65:476–82.
57. Thorell B. Flow cytometric analysis of cellular endogenous fluorescence simultaneously with emission from exogenous fluorochromes, light scatter and absorption. Cytometry. 1981;2:39–43.
58. Alturkistany F, Nichani K, Houston KD, Houston JP. Fluorescence lifetime shifts of NAD(P)H during apoptosis measured by time-resolved flow cytometry. Cytometry A. 2019;95:70–9.
59. Thorell B. Flow-cytometric monitoring of intracellular flavins simultaneously with NAD (P) H levels. Cytometry. 1983;4:61–5.
60. Niewold P, Ashhurst TM, Smith AL, King NJC. Evaluating spectral cytometry for immune profiling in viral disease. Cytometry A. 2020; 97:1165–79.
61. Diermeier-Daucher S, Brockhoff G. Flow Cytometric FRET analysis of ErbB receptor tyrosine kinase interaction. Curr Protoc Cytom. 2008; 45:12.14.1–12.14.19.
62. Szabó Á, Horváth G, Szöllősi J, Nagy P. Quantitative characterization of the large-scale association of ErbB1 and ErbB2 by flow cytometric homo-FRET measurements. Biophys J. 2008;95:2086–96.

63. Ujlaky-Nagy L, Nagy P, Szöllősi J, Vereb G. Flow Cytometric FRET analysis of protein interactions. In: Hawley TS, Hawley RG, editors. Flow cytometry protocols. New York, NY: Springer; 2018. p. 393–419.
64. Doucette J, Zhao Z, Geyer RJ, Barra MM, Balunas MJ, Zweifach A. Flow cytometry enables multiplexed measurements of genetically encoded intramolecular FRET sensors suitable for screening. J Biomol Screen. 2016;21:535–47.
65. Henderson J, Havranek O, Ma MCJ, Herman V, Kupcova K, Chrbolkova T, et al. Detecting Förster resonance energy transfer in living cells by conventional and spectral flow cytometry. Cytometry A. 2021;1–17.
66. Wu X, Simone J, Hewgill D, Siegel R, Lipsky PE, He L. Measurement of two caspase activities simultaneously in living cells by a novel dual FRET fluorescent indicator probe. Cytometry A. 2006;69:477–86.
67. Nolan JP. Flow cytometry of extracellular vesicles: potential, pitfalls, and prospects. Curr Protoc Cytom. 2015;73:13.14.1–13.14.16.

Received: 1 December 2021 | Revised: 28 June 2022 | Accepted: 4 July 2022

DOI: 10.1002/cyto.a.24669

REVIEW ARTICLE

ISAC CYTOMETRY Journal of Quantitative Cell Science PART A

A comparison of spectral unmixing to conventional compensation for the calculation of fluorochrome abundances from flow cytometric data

David Novo

De Novo Software, Pasadena, California, USA

Correspondence
David Novo, De Novo Software, Pasadena, CA 91107, USA.
Email: david.novo@denovosoftware.com

Abstract

Traditionally, flow cytometers acquired data using the same number of detectors as fluorochromes being measured in the experiment. More recently, spectral flow cytometers utilize a larger number of detectors than fluorochromes. This seemingly small difference opens the door to a wide variety of mathematical tools for the calculation of the true fluorochrome abundances from the raw detector values as compared with traditional compensation. This review will provide a brief overview of the mathematics and theory underlying traditional compensation and unmixing focusing on the differences between them and the additional information provided by unmixing approaches.

KEYWORDS
compensation, regression, spectral flow cytometry, unmixing

1 | INTRODUCTION

A flow cytometer (FC) is unique in its capability of quantitatively measuring fluorescence signals from multiple fluorochromes bound to an individual particle in an extremely short amount of time (<10 μS). This allows the cytometer to acquire an information-rich data set from many biological molecules of interest over a large number of particles, leading to robust statistics and the ability to recognize subtle differences in within cell populations in an experimental protocol [1].

The general optical configuration of FCs has not significantly changed in many decades. The FC traditionally measures fluorescence by utilizing several discrete, highly sensitive photodetectors, primarily photomultiplier tubes and/or photodiodes. A combination of dichroic and bandpass filters are arranged in order to limit specific detectors to collect photons from particular wavelengths. The filters are typically selected to correspond to the emission peaks of the different fluorochromes that are being used in the experiment, with the conceptual idea that a particular detector is the "primary" detector for a particular fluorochrome. Hence the common colloquialism of referring to detectors by the dye they are primarily designed to detect, that is, having a "FITC detector" or "PE detector," and so forth. Although the FC may have more detectors than the number of fluorochromes used in a given experiment, the instrument was almost always configured such that there were the same number of detectors active as the number of fluorochromes being measured.

The problem with this instrument configuration is well described in the literature [1–3]. In short, due to the wide emission spectrum of many fluorochromes, photons from a single fluorochrome are detected in multiple detectors. Thus, it is usually difficult to use the raw measurement from any individual detector as an indication of the abundance of any particular fluorochrome. The photons emitted from a fluorochrome that are detected in the "non-primary" detectors for that fluorochrome are said to be "spilling over" into these other detectors, with the conceptual framework being that these photons are going where they should not be. A correction process, termed compensation, was devised to either eliminate these "spilled over" photons and/or "return them" to the detector in which they were designed to be detected.

For clarity, the standard notation is followed for the equations within this paper: lower case and bolded variables represent vectors; upper case and bolded variables represent a matrix and scalars are lower case, and italicized.

1.1 | Theory and mathematics of compensation

In early FCs, compensation was performed using analog circuit that would subtract a fraction of the signal that was measured in one detector from another detector [1, 4]. The precise fraction to be subtracted was manually determined by the user by physically adjusting controls (dials or buttons) on the cytometer itself, according to certain heuristics. In 1993, Bagwell and Adams proposed a model that could mathematically calculate the abundance of an arbitrary number of fluorochromes across the same number of detectors [5]. The approach has significant advantages compared with the previous manual approaches and involved obtaining data from single-stained controls and using the mean value from each control in each detector to construct a "spillover matrix." A relationship was assumed such that

$$r = Ma_T \tag{1}$$

where **r** represents the observed raw detector values from an arbitrary particle, **M** is the spillover matrix for that particle and $\mathbf{a_T}$ is the true abundance of each fluorochrome on that particle. Since it is impossible to measure **M** for each experimental particle, an average mixing matrix ($\mathbf{M_{Avg}}$) was experimentally determined. Since **r** is a measured value, and both **M** and $\mathbf{M_{Avg}}$ are square matrices, it becomes possible to solve for $\mathbf{a_T}$

$$a_C = M_{Avg}^{-1} r \tag{2}$$

where $\mathbf{M_{Avg}^{-1}}$(often called the compensation matrix) is the inverse of $\mathbf{M_{Avg}}$, $\mathbf{a_C}$ is the calculated abundance of each fluorochrome on the particle. This model has been used, essentially unchanged, since the initial publication. We will be exploring the relationship between the calculated abundance ($\mathbf{a_C}$) and true abundance ($\mathbf{a_T}$) throughout this article. (In an ideal world $\mathbf{a_C}$ from Equation (2) would equal $\mathbf{a_T}$ from Equation (1), however, as described below, this is rarely the case, and the notation for distinguishing between calculated and true abundances is introduced to reflect that.)

1.2 | Theory and mathematics of unmixing

Independently, other scientific fields, such as geology [6–8], remote sensing [9, 10], and chemometrics [11] were encountering similar problems, whereby the measurements arose from a mixture of a potentially unknown number and abundance of different elements termed endmembers. From the beginning, the instruments in these other fields were configured in such a way that the number of detectors (often vastly) exceeded the number of endmembers. Because of this, there was never the notion that any particular detector was supposed to be the "primary" detector, nor that the end member photons were "spilling" anywhere they were not supposed to be, which was the prevalent conceptual framework underlying traditional compensation [3]. Instead, the mathematics of mixture modeling was utilized, whereby each detector is assumed to be detecting a mixture of photons from multiple endmembers, and the mathematics was utilized to "unmix" the data and calculate the actual abundance of each end member [12]. The matrix representing the pure spectra of the individual endmembers was termed the "mixing" matrix, instead of spillover matrix, but the same basic model as shown in Equation (1) was used. The key difference is that **M** and $\mathbf{M_{Avg}}$ are not square, but rectangular, since there are more detectors than endmembers.

More recently, "spectral" FCs (SFC) measuring fluorescence [13] and Raman scatter [14] have been built in the laboratory [15] and are now commercially available from several instrument vendors. A detailed description of their optical and detection schematics is beyond the scope of this review; however, they all acquire data using more detectors than there are endmembers in the experiment and unmix the data using unmixing as opposed to compensation.

1.3 | Contribution of noise to calculated abundances

There are two important sources of noise in a FC measurement. The one most familiar to cytometrists is instrument noise (due to noise in electronics, stray light, etc.) which will be termed ε_O. The other noise (termed ε_M) is the difference between the actual emission from an individual particle (**M**) and $\mathbf{M_{Avg}}$ that is used in the unmixing mathematics. It is ε_M that is generally the largest contributor to the "spread" commonly seen in compensated/unmixed data.

It is important to note that construction of $\mathbf{M_{Avg}}$ is similar for both compensation and unmixing. In general, $\mathbf{M_{Avg}}$ is obtained by running large numbers of single stained controls (cells or beads) and calculating a mean or median in each detector. After normalization, this results in an average emission spectrum ($\mathbf{em_{Avg}}$) for each end member as measured by the instrument for a particular set of detector gains and optical configuration. By convention, the columns of $\mathbf{M_{Avg}}$ represent the $\mathbf{em_{Avg}}$ of each end member, and the rows represent a particular detector. It is important to note that it is practically impossible for **M** of each particle to equal $\mathbf{M_{avg}}$. Photon emission is a random process, governed by Poisson statistics. The short dwell time of the cells in the laser can result in relatively few photons being emitted and for the emission profile of individual fluorochromes to deviate significantly from average (Figure 1).

Thus, we can define

$$M_{Diff} = M_{Avg} - M \tag{3}$$

and reformulate Equation (1)

$$\mathbf{r} = M_{Avg}\mathbf{a}_T + M_{Diff}\mathbf{a}_T \tag{4}$$

we can define ε_M

$$\varepsilon_M = M_{Diff}\mathbf{a}_T \tag{5}$$

as the portion of **r** resulting from the random differences between the photons emitted by a particular particle and $\mathbf{M_{Avg}}$. This is a random

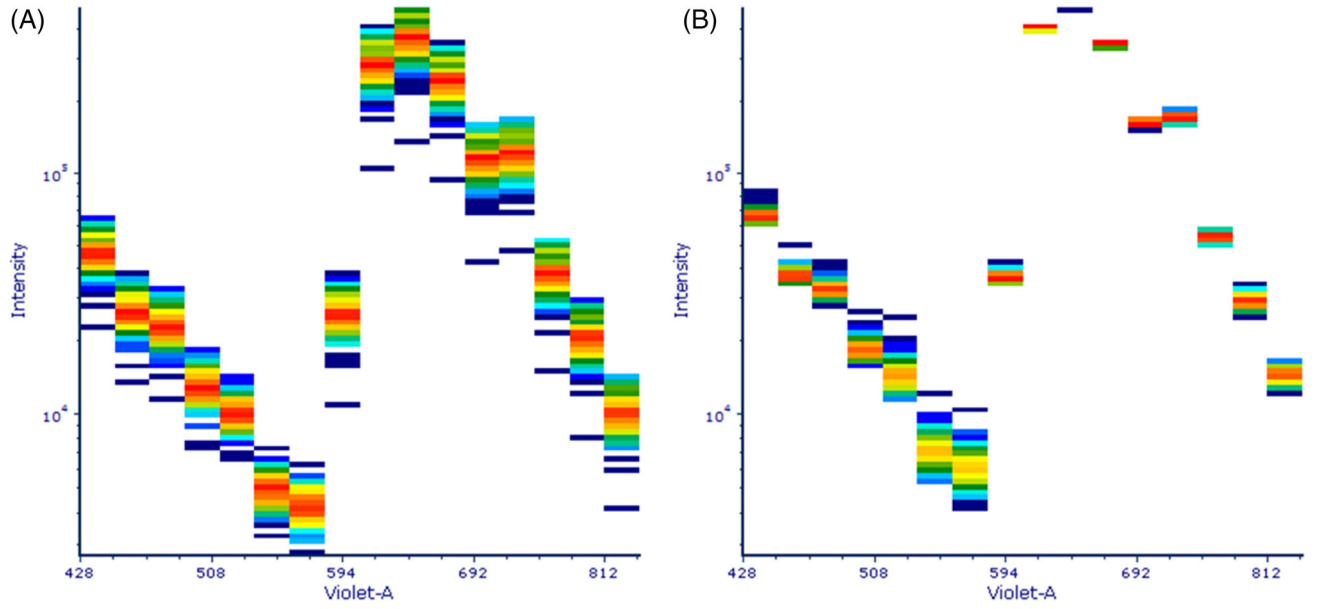


FIGURE 1 *Differences in emission spectra from multiple bead particles*. Data from beads stained with BV-605 were acquired on a Cytek Aurora flow cytometer and gated on singlets and bright staining. Panel A—the raw spectral data of the gated population. There is a wide distribution in emissions across the beads. Panel B—the emission were normalized such that the peak of the spectra were identical across all beads. This should normalize for any differences in the amount of dye present in different beads. In spite of this, there is still a wide variation in the emission profile across different beads [Color figure can be viewed at wileyonlinelibrary.com]

value, it is termed "noise," similar to any other random signal measured by a detection system. Defining ε_O as any other noise (instrument, electrical, light, etc.). A more complete reformulation of Equation (1) becomes

$$\mathbf{r} = \mathbf{M}_{Avg}\mathbf{a}_T + \varepsilon_M + \varepsilon_O \tag{6}$$

Modern commercial flow cytometers have a highly sophisticated design such that ε_O is often negligible for a wide variety of experimental conditions, the notable exception being the measurement of small or dim particles. However, aside from collecting a dramatically larger number of photons per fluorochrome (which is at odds with the other design parameter of collecting data from large numbers of cells in short period of time) no technological solution will be able to eliminate ε_M since it is intrinsic to the photon emission process itself.

The presence of ε_M that follows a Poisson distribution causes the standard linear models used in unmixing/compensation to become inappropriate, as shall be discussed subsequently. It is now simple to understand how ε_M contributes to "spreading" artifacts often associated with traditional compensation. Ignoring ε_O and rearranging Equation (6) yields

$$(\boldsymbol{r} - \varepsilon_M)\boldsymbol{M}_{Avg}^{-1} = \boldsymbol{a}_C \tag{7}$$

Comparing Equations (7) to (1), it is immediately obvious that $\mathbf{a}_C$ only equals $\mathbf{a}_T$ in the special case that $\boldsymbol{\varepsilon}_M = 0$ (i.e., all endmembers on the individual particle emitted exactly according to their $\mathbf{em}_{Avg}$), which rarely happens. In all other cases, the values from ε_M are spread throughout the $\mathbf{a}_C$ according to $\boldsymbol{M}_{Avg}^{-1}$ such that $\mathbf{a}_C$ can differ quite significantly from the true abundances that were present on the particle.

It is important to note that Equations (2) and (7) are only valid solutions when $\mathbf{M}_{Avg}$ is a square matrix. When $\mathbf{M}_{Avg}$ is rectangular, $\mathbf{M}_{Avg}{}^{-1}$ cannot be mathematically calculated, and the solution to Equation (1) becomes [16, 17].

$$\boldsymbol{a}_C = \left(\boldsymbol{M}_{Avg}^T\boldsymbol{M}_{Avg}\right)^{-1}\boldsymbol{M}_{Avg}^T\boldsymbol{r} \tag{8}$$

Where $\boldsymbol{M}_{Avg}^T$ is the transpose of $\mathbf{M}_{Avg}$. It can easily be shown that when $\mathbf{M}_{Avg}$ is a square matrix, Equation (8) simplifies to Equation (2), highlighting that compensation is special case of a more generalized mixture model.

2 | BENEFITS OF UNMIXING

Conceptually, the difference between a spectral and conventional FC is minimal. The distinction can boil down to whether one acquires (and mathematically utilizes) measurements from a larger number of detectors than one has endmembers. However, this seemingly small distinction affords spectral FC many advantages from a data analysis point of view. In theory, spectral cytometry can collect more photons from an experiment, compared with conventional flow cytometry, since there are generally less gaps between the filters in a spectral cytometer. More photons should result in more accurate measurements and hence better data. However, it is unclear how critical this benefit is in practice, since typically enough photons are acquired in

either case to obtain data sufficient for the required experimental precision.

The main benefit (from a data analysis point of view) of having more detectors than endmembers is that it results in an overdetermined system of equations, which is defined as a system where there are more equations than unknowns. This is reflected by a rectangular **M**. There are two main benefits of an overdetermined system of equations are: the ability to obtain a measurement of the noise and to improve the condition number.

2.1 | Additional detectors acting as replicate measurements

It is an elementary tenet of sampling theory that if one obtains a measurement contaminated by noise it would be unreasonable to presume the true underlying value is the same as the measured value. To improve the accuracy of the estimation of the true underlying value one takes multiple replicate noisy measurements. By understanding the noise model one can combine these measurements together in some way (often by simply taking an average) to obtain an accurate estimate of the true underlying value, even though the individual measurements can be inaccurate. In general, the more replicates one performs, the better the estimation of the underlying value.

A system with more detectors than endmembers is effectively taking multiple replicates of the same sample. These replicates contain information regarding both the underlying endmember abundance and the noise in the instrument. Consider the case of a particle stained with a single dye, with a known $\mathbf{em_{Avg}}$, measured across multiple detectors. In the absence of any noise, the additional detectors are irrelevant since the value in any detector can be predicted from the value in any other detector, simply by applying $\mathbf{em_{Avg}}$. However, if there is noise, and the noise is independent across the detectors, the additional detectors provide information about the noise, which can then be treated appropriately to provide a better estimate of the underlying abundance of the single stained control. As above, the more replicates (i.e., additional detectors) the better the improvement one can make in the abundance estimation.

2.2 | Improvements in the condition number of M

The condition number of the matrix quantifies how changes in the inputs (**r**) can affect the output ($\mathbf{a_C}$). Starting from Equation (1) one can define

$$\boldsymbol{r}+\boldsymbol{\varepsilon r}=\boldsymbol{M}(\boldsymbol{a_C}+\boldsymbol{\varepsilon}(\boldsymbol{a_C})) \tag{9}$$

Where $\boldsymbol{\varepsilon}(\mathbf{a_C})$ is the error in $\mathbf{a_C}$ resulting from an error in **r** ($\boldsymbol{\varepsilon_r}$). The condition number of **M**, $\kappa(\mathbf{M})$ is defined as

$$\frac{\|\epsilon r\|}{\|r\|} \leq \kappa(\boldsymbol{M})\frac{\|\epsilon a_C\|}{\|a_C\|} \tag{10}$$

$\kappa(\mathbf{M})$ acts as a factor that (if $\kappa(\mathbf{M})>1$) that provides an upper bound on the magnification of the relative change in **r** ($\frac{\|\epsilon r\|}{\|r\|}$) into a relative change in $\mathbf{a_C}$ ($\frac{\|\epsilon a_C\|}{\|a_C\|}$).

Intuitively, very small changes in **r** should result in small changes in $\mathbf{a_C}$. However, if $\kappa(\mathbf{M})$ is large then small changes in **r** may propagate into large changes in $\mathbf{a_C}$. In the case of FC, we have a single opportunity to measure the *r* of a particular particle, we which we know is contaminated with $\boldsymbol{\varepsilon_M}$. If $\kappa(\mathbf{M})$ is large, $\boldsymbol{\varepsilon_M}$ will highly magnified and result in a large change in $\mathbf{a_C}$. Abundance estimations from matrices with a smaller $\kappa(\mathbf{M})$ will be less affected by $\boldsymbol{\varepsilon_M}$, no matter what noise model is used in the minimization process.

Presuming the columns of a matrix are linearly independent, the condition number is highly influenced by the differences in the columns of the matrix, that is, the more the columns of the matrix are different from each other, the lower the condition number. Figure 2 shows results from simulations which considers a series of FC measurements of identical particles stained with two dyes. Since the particles themselves are identical the only difference between the measurements is $\boldsymbol{\varepsilon_M}$ which was simulated as having a Gaussian distribution. Figure 2A shows **M** with a condition number of 162 from a two dye/two detector system where the dyes are very similar to each other. Figure 2B shows the simulation calculations performed on a representative particle using the matrix in Figure 2A. Figure 2B-Line 1 shows then when multiplying a particle with known abundance [200, 100] by the matrix in Figure 2A, the observation **r** [190, 189] is obtained. Figure 2B-Line 2 shows that when multiplying **r** [190,189] by the inverse of the matrix in Figure 2A, the original abundance [200, 100] is accurately recovered. Figure 2B-Line 3 shows that even if a small amount of noise is added to **r** (compare the first term of line 2 and line 3), and then multiplied by the same inverse, a dramatically different abundance is calculated. Figure 2C–E show different matrices used for the simulations (see figure legend for rationale for matrix selection). Figure 2F shows the normalized calculated abundance for dye 1 from simulations using the matrices from Figure 2A, C–E. In all cases, the mean recovered abundance, and the peak of the distributions, is at the true value of 200. However, except for the simulation using the matrix 2E, there was a huge variation in the recovered abundance values. This means that on any given sampling of our simulated particle (and similarly obtaining our single measuring of the particle in the FC), we are likely to calculate an abundance that is quite different from the true one. Similar results were obtained when unmixing the second dye, and results are summarized in Figure 2G.

Of interest to note is that the matrix in Figure 2C is simply the matrix from Figure 2A with each detector split equally into two detectors. The condition numbers of matrices in Figure 2A,C are identical, as expected. However, Figure 2F shows that the *SD* of the calculated abundances is lower using the matrix from Figure 2C compared to the one from Figure 2A. This is because the matrix from Figure 2C effectively provides replicated measurements (compared with 2A) which can be used to average out the noise. The matrix in Figure 2D is very slightly different from Figure 2C, however, even this slight difference results in a significantly lower condition number (Figure 2G) and much lower *SD* of $\mathbf{a_C}$. Unsurprisingly, the matrix from Figure 2E results in

(A)

Dye 1	Dye2
0.9	0.89
0.1	0.11

(B)

1 $[200 \quad 100]\begin{bmatrix}0.9 & 0.89\\ 0.1 & 0.11\end{bmatrix} = [190 \quad 189]$

2 $[190 \quad 189]\begin{bmatrix}11 & -89\\ -10 & 90\end{bmatrix} = [200 \quad 100]$

3 $[189 \quad 191]\begin{bmatrix}11 & -89\\ -10 & 90\end{bmatrix} = [169 \quad 369]$

(C)

Dye 1	Dye2
0.45	0.445
0.45	0.445
0.05	0.055
0.05	0.055

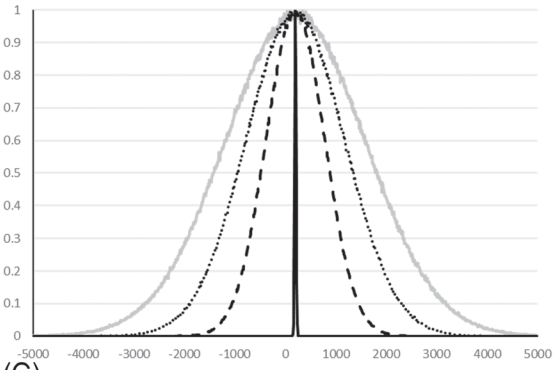
(D)

Dye 1	Dye2
0.45	0.45
0.45	0.44
0.05	0.06
0.05	0.05

(E)

Dye 1	Dye2
0.6	0.39
0.3	0.5
0.09	0.04
0.01	0.07

(F)



(G)

Matrix	Condition #	Dye 1 mean	Dye 1 SD	Dye 2 mean	Dye 2 SD
A	162.4	201.1	1391	98.8	1404
C	162.4	198.5	983	101.4	993
D	99.9	199.9	585	100.1	590
E	4.4	200	15	100	16

FIGURE 2 *The effects of condition number on the estimation of recovered abundances*. Panel A—Hypothetical M from a 2 dye/2 detector system. Columns are normalized to 1. Panel B algorithmic steps involved in simulating a single sample with noise (details in text). Panel C, a mixing matrix where the detectors from 2A are split equally in two. Panel D, a mixing matrix with very slight differences from Panel C. Panel E—mixing matrix with significantly different columns. Note, all matrices have the same integrated emission in the top and bottom half compared to each other. Panel F—the distribution of the calculated abundance of dye 1 from 4,000,000 simulations of each matrix (the number of trials required to get the SEM of the results from matrix A to ~1). Gray solid line—Matrix 2A, black dotted line—Matrix 2C, black dashed line—Matrix 2D, black solid line—Matrix 2E. Panel 2G—A summary of the calculated abundances for the two dyes using the different matrices. Note that the means are all similar, but the *SD* varies dramatically based on the matrix. For each simulation, the known abundance [200,100] was multiplied by the respective mixing matrix to obtain a calculated observation. Noise was added to each detector in the observation. The noise was drawn from a Gaussian distribution with a mean of 0 and *SD* equal to half of the smallest value of the calculated observation in order to generate homoscedastic data. The data were then unmixed using Ordinary Least Squares to obtain the calculated abundances for that trial.

the lowest condition number and smallest *SD* of the recovered abundances since the columns are dissimilar from each other and the condition number is much lower than the other matrices.

3 | WILL UNMIXING ALWAYS IMPROVE YOUR ABUNDANCE ESTIMATION?

While even adding a single extra detector will allow the calculation of $\mathbf{a_C}$ using Equation (8) and/or sophisticated matrix algebra techniques, it does not necessarily guarantee significantly improved results (i.e., where $\mathbf{a_C}$ is closer to $\mathbf{a_T}$). The results will only improve if the extra detectors provide an improved condition number and/or sufficient replicate measurements to meaningfully average out the noise. This is shown in Figure 2, where only if the detectors in Figure 2A were split in such a way that the columns became different (compare Figure 2C, D with 2E) do the results dramatically improve (Figure 2G). The main advantage of a spectral system in this regard is the ability to exploit the differences in the multiple endmember emission spectra (and hence lower the condition number) without performing manual tuning of the filters for each experimental panel.

We have already established that since ε_M is not accounted for in Equation (2) or (8) in almost all cases $\mathbf{a_C} \neq \mathbf{a_T}$. In the case of compensation (since **M** is square) there is a unique solution for $\mathbf{a_C}$, i.e. if you replace $\mathbf{a_T}$ in Equation (1) with $\mathbf{a_C}$ calculated from Equation (2), the result that is calculated ($\mathbf{r_C}$) will match the actual observation **r**. When **M** is rectangular, the system of equations becomes overdetermined, in almost all situations if you take $\mathbf{a_C}$ calculated from Equation (8) and substitute it into $\mathbf{a_T}$ in Equation (1) then $\mathbf{r_C} \neq \mathbf{r}$. This leads to the seemingly strange situation where not only do we know that $\mathbf{a_C} \neq \mathbf{a_T}$, but we also cannot even use that value of $\mathbf{a_C}$ to accurately derive the value of **r** which we initially used to calculate $\mathbf{a_C}$. This is a fundamental property of overdetermined systems of equations and is not as counterintuitive as it may seem. Scientists are used to the fact that when performing a linear regression, the regression line never goes through every data point that was used to calculate the slope and intercept (unless the data points are exactly on a straight line). The regression cannot exactly recapitulate the observed data and is considered a "best fit" to the data. When determining the slope and intercept of this "best fit" line, one typically which minimizes the Euclidian distance between the calculated fit and the individual data points that were used to generate the fit.

4 | ROLE OF PRESUMED NOISE MODEL IN UNMIXING CALCULATIONS

If one presumes Gaussian noise, the solution for $\mathbf{a_C}$ in Equation (8) is actually an analytical solution to the "best fit" or maximum likelihood estimation (MLE) of $\mathbf{a_T}$, that is, when substituting $\mathbf{a_T}$ in Equation (1) with $\mathbf{a_C}$, the calculated $\mathbf{r_C}$ will be the closest to **r** compared to all other values of $\mathbf{a_C}$. Hence Equation (8) is often termed the Ordinary Least Squares (OLS) solution. In general, determining the MLE is an iterative process that conceptually proceeds as follows:

1. Guess an initial value for $\mathbf{a_C}$
2. Calculate $\mathbf{r_C}$,

3. Calculate the distance between $\mathbf{r}_c$ and $\mathbf{r}$ (call this distance $\mathbf{d}$)
4. Make another guess of $\mathbf{a}_C$ based on an approach designed to minimize $\mathbf{d}$
5. Calculate the new $\mathbf{r}_C$, and recalculate $\mathbf{d}$
6. if the change in $\mathbf{r}_C$ is below some threshold, return to step 4, otherwise we have found the final $\mathbf{a}_C$

As can be appreciated from the algorithm above, the resulting $\mathbf{a}_C$ is highly dependent on how one calculates $\mathbf{d}$ (and to a lesser extent the initial guess of $\mathbf{a}_C$). Equation (8) will result in $\mathbf{a}_C$ that is the MLE of $\mathbf{a}_T$ under the assumption that $\mathbf{d}$ is L_2 norm, or Euclidian distance, which presumes that the mean of the noise is zero, and the *SD* of the noise is the same in each detector. It is extremely convenient to use Equation (8) since $\mathbf{a}_C$ can be calculated quite rapidly without an iterative process. In fact, since compensation is simply a special case of unmixing (i.e., Equation (2) is a special case of Equation (8)), compensation results can also be thought of as an MLE based on an assumption of Gaussian noise. If $\boldsymbol{\varepsilon}_M$ does follow a Gaussian distribution then $\mathbf{a}_C$ will be the best approximation we can derive for $\mathbf{a}_T$. However, if $\boldsymbol{\varepsilon}_M$ does not follow a Gaussian distribution then not only will $\mathbf{a}_C \neq \mathbf{a}_T$ (which will almost always be the case) $\mathbf{a}_C$ will not even be the MLE for $\mathbf{a}_T$ since the presumed noise model is not even correct [18].

In reality $\boldsymbol{\varepsilon}_M$ is expected to follow a Poisson distribution due to the fundamental stochastic process of photon emission [1, 18]. The hallmark of the Poisson is that it is heteroskedastic; meaning that the *SD* of the noise is related to the intensity of the signal. That is, detectors with more signal will have more absolute noise than detectors with less signal. Calculating $\mathbf{a}_C$ using a model that assumes a Gaussian noise when $\boldsymbol{\varepsilon}_M$ that follows a Poisson distribution will introduce a particular bias into $\mathbf{a}_C$. Detectors with large signal values will be overfit, that is, have more weight in the fitting process, than they should. This will cause the MLE fitting process to force $\mathbf{r}_C$ to match $\mathbf{r}$ as closely as possible for the detectors with large signals at the expense of letting $\mathbf{r}_C$ in detectors with dimmer signals deviate from the corresponding values in $\mathbf{r}$. This is because the distance minimization expects that the amount of noise in the bright and dim detectors is identical, when in fact they are not. Hence over a large population of cells, the low abundance endmembers (which generate dim detector signals) have a broader distribution than they should since the Gaussian MLE does not put sufficient weight on signals from the dimmer detectors. In contrast, an algorithm that minimizes a distance metric based on the assumption of Poisson noise (i.e., the Kulback-Leibler divergence) will place more emphasis on fitting dim detectors and decrease the *SD* of low abundance endmembers (at the expense of a slightly increased distribution of the bright endmembers). There are both analytical (weighted least squares [WLS]) and iterative (iteratively reweighted least squares [IRLS]) solvers that can calculate an MLE that attempts to minimize a KL divergence [20]. The mismatch between the actual noise model and the model assumed by the unmixing can further exacerbate the spillover spreading that is often observed in flow cytometry. This has been previously shown both theoretically and in practice [18].

5 | APPLICATIONS TAKING ADVANTAGE OF THE OVERDETERMINED NATURE OF UNMIXING

Another benefit of SFCs is that the unmixing algorithms can often accurately unmix data from fluorophores with similar emission spectra without custom tuning of the filters by the user [19, 20]. By minimizing D, the equations naturally utilize key detectors which results in separation of the various endmembers in a way that would be difficult to predict a priori, particularly when using a Poisson based unmixing model.

Another area where spectral cytometry can have a significant impact is in the treatment of autofluorescence. Since conventional compensation only utilizes the same number of detectors as fluorochromes, there was no way to obtain additional information about the autofluorescence of the cells. Conventional approaches are to simply assume the autofluorescence is similar on all cells, or assume that the autofluorescence is negligible and ignore it [2]. Due to the additional detectors in a SFC it is possible to simply treat the autofluorescence as another endmember in the system, and unmix it explicitly along with the fluorochromes. Thus, the abundance of autofluorescence per cell will be calculated along with the fluorochromes. In fact, different autofluorescence species with different spectral signatures in individual or distinct cell types can easily be calculated, which is simply not possible with conventional flow cytometry using a square mixing matrix.

Another benefit of spectral cytometry, from a data analysis point of view, is the ability to use many of the advanced mathematical techniques that were developed for the analysis of overdetermined systems in other fields. For example, blind unmixing attempts have evolved over the past 30 years [21–23] to determine the mixing matrix as well as the per-event endmember abundances from the experimental data, without the need for single stained controls. This was recently applied to spectral cytometry data [24] and may be useful in cases where there is insufficient sample to obtain ideal single stained controls.

6 | CONCLUSION

FC utilizing conventional compensation has yielded a wealth of important results that has impacted almost all facets of biology. However, the application of spectral unmixing to FC data affords new possibilities that are simply not possible utilizing traditional compensation. As can be seen from many of the examples above, much of the advantage of spectral cytometry stems from the overdetermined nature of the problem. The fact that there is no unique solution allows the data analysis to apply different mathematical models that will result in different solutions. As with all mathematical modeling, the relevance of the obtained results will be dependent on how closely the model mirrors the actual underlying physical processes being studied. Spectral flow cytometry allows one to both apply a wide variety of models to the data and evaluate the goodness of fit using standard regression

techniques. In addition, the conceptual shift from "correcting spillover error" that has historically been prevalent with compensation, to one of applying a variety mathematical models to overdetermined data can result in a more sophisticated and beneficial treatments of cytometric data.

AUTHOR CONTRIBUTIONS

David Novo: Conceptualization (equal); software (equal); writing – original draft (equal).

ACKNOWLEDGMENTS

I would like thank Dr. Bartek Rajwa for his many stimulating discussions on this topic and for providing the historical references contained herein. I would also like to thank Joanne Lannigan for her many comments.

CONFLICT OF INTEREST

I have no conflicts of interest to declare.

ORCID

David Novo https://orcid.org/0000-0003-1713-0605

REFERENCES

1. Shapiro HM. Practical flow cytometry. 4th ed. New York: Wiley-Liss; 2003. p. 681.
2. Roederer M. Compensation in flow cytometry. Curr Protoc Cytom. 2002;22:Unit 1.14.
3. Roederer M. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. Cytometry. 2001;45:194–205.
4. Loken MR, Parks DR, Herzenberg LA. Two-color immunofluorescence using a fluorescence-activated cell sorter. J Histochem Cytochem. 1977;25:899–907.
5. Bagwell CB, Adams EG. Fluorescence spectral overlap compensation for any number of flow cytometry parameters. Ann N Y Acad Sci. 1993;677:167–84.
6. Full WE, Ehrlich R, Klovan JE. EXTENDED QMODEL? Objective definition of external endmembers in the analysis of mixtures. J Int Assoc Math Geol. 1981;13:331–44.
7. Imbrie J, Van Andel TH. Vector analysis of heavy-mineral data. Geol Soc Am Bull. 1964;75:1131.
8. Dymond J. Geochemistry of Nazca plate surface sediments: an evaluation of hydrothermal, biogenic, detrital, and hydrogenous sources. Vol 154. Boulder CO: Geological Society of America; 1981. p. 133–74.
9. Horwitz H, Hyde P, Richardson W. Improvements in estimating proportions of objects from multispectral data. NASA. 1975;75:1-75.
10. Keshava N, Mustard JF. Spectral unmixing. IEEE Signal Process Mag. 2002;19:44–57.
11. Lawton WH, Sylvestre EA. Self modeling curve resolution. Dent Tech. 1971;13:617–33.
12. Renner RM. The resolution of a compositional data set into mixtures of fixed source compositions. Appl Stat. 1993;42:615.
13. Grégori G, Patsekin V, Rajwa B, Jones J, Ragheb K, Holdman C, et al. Hyperspectral cytometry at the single-cell level using a 32-channel photodetector. Cytometry A. 2012;81A:35–44.
14. Nolan JP, Condello D, Duggan E, Naivar M, Novo D. Visible and near infrared fluorescence spectral flow cytometry. Cytometry A. 2013; 83A:253–64.
15. Nolan JP, Condello D. Spectral flow cytometry. Curr Protoc Cytom. 2013;63:Unit 1.27.
16. Penrose R. A generalized inverse for matrices. Math Proc Camb Philos Soc. 1955;51:406–13.
17. Dresden A. The fourteenth western meeting of the American Mathematical Society. Bull Am Math Soc. 1920;26:385–97.
18. Novo D, Grégori G, Rajwa B. Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices. Cytometry A. 2013;83A:508–20.
19. Futamura K, Sekino M, Hata A, Ikebuchi R, Nakanishi Y, Egawa G, et al. Novel full-spectral flow cytometry with multiple spectrally-adjacent fluorescent proteins and fluorochromes and visualization of in vivo cellular movement. Cytometry A. 2015;87:830–42.
20. Ferrer-Font L, Small SJ, Lewer B, Pilkington KR, Johnston LK, Park LM, et al. Panel optimization for high-dimensional Immunophenotyping assays using full-spectrum flow cytometry. Curr Protoc. 2021;1:e222.
21. Sasaki K, Kawata S, Minami S. Constrained nonlinear method for estimating component spectra from multicomponent mixtures. Appl Opt. 1983;22:3599–603.
22. Neher RA, Mitkovski M, Kirchhoff F, Neher E, Theis FJ, Zeug A. Blind source separation techniques for the decomposition of multiply labeled fluorescence images. Biophys J. 2009;96:3791–800.
23. Gobinet C, Perrin E, Huez R. Application of non-negative matrix factorization to fluorescence spsctroscopy, 4.
24. Jiménez-Sánchez D, Ariz M, Morgado JM, Cortés-Domínguez I, Ortiz-de-Solórzano C. NMF-RI: blind spectral unmixing of highly mixed multispectral flow and image cytometry data murphy R, editor. Bioinformatics. 2020;36:1590–8.

Received: 16 September 2021 | Revised: 25 March 2022 | Accepted: 1 April 2022

DOI: 10.1002/cyto.a.24557

ORIGINAL ARTICLE

MIFlowCyt

ISAC INTERNATIONAL SOCIETY FOR ADVANCEMENT OF CYTOMETRY

CYTOMETRY Journal of Quantitative Cell Science PART A

Building a spectral cytometry toolbox: Coupling fluorescent proteins and antibodies to microspheres

Simon Monard

WEHI, Flow Cytometry (FACS) Facility, Parkville, Australia

Correspondence
Simon Monard, BSc Hons, MSc, Head of Flow Cytometry, Walter and Eliza Hall Institute, 1G Royal Parade, Parkville, Victoria 3052, Australia.
Email: monard@wehi.edu.au

Abstract

Fluorescent proteins (FPs) have become an essential tool for biological research. Since the isolation and description of GFP, hundreds of fluorescent proteins have been discovered and created with various characteristics. The excitation of these proteins ranges from ultra-violet (UV) up to near infra-RED (NIR). Using conventional cytometry with each detector assigned to each fluorochrome, great care must be taken when selecting the optimal bandpass filters to minimalize the spectral overlap. In the last 8 years, several companies have released full spectrum flow cytometers which eliminates the need to change optical filters for analyzing FPs. This addressed at least part of the problem however, the laser wavelengths in commercial instruments are generally not ideal for all fluorescent proteins yet do allow the separation of at least six FPs. Another technical challenge is to have convenient single color controls. If four different FPs are being used in an experiment, single color controls will be needed to compensate or unmix the data. In the case of cultured cells this will involve having each of the FPs expressed in cell lines separately with a parental cell line expressing none. In the case of in vivo experiments, colonies of animals may need to be maintained expressing each FP along with a wildtype animal. This represents a considerable expense and inconvenience. An appealing alternative is to produce and purify FPs and covalently couple to polystyrene microspheres. Such microspheres are ready to use and can be stored at 4°C for months or even years without any deterioration in fluorescence. The same procedure can be used to couple antibodies to these particles. Here we describe this procedure which can be executed in any lab without any special equipment or skills.

KEYWORDS

fluorescent proteins, full spectrum cytometry, microspheres, reference controls, spectral unmixing

1 | INTRODUCTION

Since the discovery, isolation and cloning of GFP [1–3], hundreds of fluorescent proteins have been discovered and created. A valuable resource to explore all of these proteins is the FP database [4]. Fluorescent proteins are widely used in flow cytometry and imaging applications, from simple experiments where an FP is used to demonstrate the successful transfection or transduction of cells to more complex experiments where several FPs are used to demonstrate cell cycle state [5, 6] or various gene activities. Fluorescent proteins are available which can be excited from ultra-violet (UV) to near infra-red (NIR) and range in their brightness. There are many excellent reviews on the applications of fluorescent proteins in biology [7–9]. The emission spectra of FPs are often quite broad which

has made the use of multiple FPs concurrently quite challenging, as is attaining single color controls for each FP. For instance, the "brainbow" mouse uses the confetti system which expresses green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP) and cyan fluorescent protein (CFP) within different cells. To analyze these cells by flow cytometry would require each of the FPs expressed separately for compensation or unmixing.

Flow cytometry is an essential tool for both biological research and clinical diagnostic services and has been used in research setting for about 50 years. Conventional cytometers have a number of optical detectors which are usually photomultiplier tubes (PMTs) but several instruments use avalanche photo diodes (APDs) and more recently Silicon Photomultipliers. When using multiple fluorescent dyes in an experiment, each dye will be intended for one detector. However, very often the fluorescence of one dye will also spill into detectors intended for other dyes. This can be corrected for by using color compensation [10, 11]. Generally, the same number of detectors are used as the number of different dyes used in the experiment with any superfluous detectors not being used in data acquisition thus reducing the size of the data file.

The first full spectrum cytometer became commercially available in 2014 when Sony released the SP6800™ instrument. It used a single 32 channel PMT as the detector. Cytek Bio released their Aurora™ full spectrum analyzer in 2017 which initially had three spatially separated lasers and arrays of APDs as detectors. The initial laser configuration was 405, 488, and 640 nm. Later a 561 nm laser and then a 355 nm UV laser was added. Sony released a more advanced instrument in 2019 the ID7000™ and Becton Dickinson started offering a full spectrum option of their flagship high end analyzer the Symphony A5™. Full spectrum cytometers, although they share many components with conventional cytometers, are used a little differently. They will have more detectors than fluorochromes and all the fluorescent detectors will be used for every experiment and will produce a spectral signature for each dye. The single-color controls are called reference controls and instead of traditional compensation, spectral unmixing is used [11]. All the work described herein has been generated using a Cytek Aurora™.

Functionalized polystyrene microspheres are available commercially from several sources. Carboxyl microspheres allow coupling to the amine terminus of any protein [10, 12, 13] however, it is not guaranteed that the function of the protein will not be compromised. Microspheres are available in a variety of sizes but throughout this study 4.5 μm microspheres were used.

2 | MATERIALS AND METHODS

2.1 | Producing and purifying fluorescent proteins

All fluorescent proteins except eGFP were expressed and purified by Monash University Protein Production facility from plasmids purchased from Addgene.org. The eGFP was a gift from David Miller and Jacqui Gulbis Structural Biology division, Walter and Eliza Hall Institute for Medical Research.

2.2 | Expression

Using the provided cultures from Addgene, the plasmid DNA was purified and transformed into LMG194 *Escherichia coli* strain with ampicillin as a selection marker. A culture from a single colony was grown in 50 ml of LB broth with ampicillin overnight at 37°C. Five milliliter of this culture was inoculated into 500 ml of LB containing ampicillin and grown until the OD reached 0.6. The protein expression was induced by the addition of 0.2% Larabinose and grown overnight at 37°C with shaking at 180 rpm. The biomass was centrifuged then the pellets were resuspended in lysis buffer and stored at −80°C until ready to process. One liter of biomass was generated for this purification.

2.3 | Purification

The cell paste was thawed and a protease inhibitor tablet and a small amount of DNase I was added to each sample. This material was sonicated then centrifuged at 20,000 *g* for 15 min at 4°C. The material was filtered and loaded onto a 1 ml nickel affinity column at 1 ml/min and washed with 100 mM Na_3PO_4, pH 7.4, 150 mM NaCl, 20 mM Imidazole, 10% glycerol. The bound protein was eluted from the column using a high imidazole elution step and the material was loaded onto a S200 gel filtration 16/60 column 100 mM Na Phosphate, pH 7.4, 150 mM NaCl, 10% glycerol. A series of fractions were analyzed on an SDS-PAGE gel. Those fractions that contained a protein of the expected size were pooled. The concentration of the final fraction was measured using a Biorad protein assay. The FPs produced were Venus, eBFP2, Cerulean, dsRED homodimer, Td Tomato, mCherry and mKate. For long term storage, purified FPs were frozen at −20°C in 50% glycerol in PBS.

3 | COUPLING ANTIBODIES AND FLUORESCENT PROTEINS TO MICROSPHERES

Proteins may be covalently coupled to carboxyl functionalized microparticles by using water-soluble carbodiimide (EDAC) to activate the carboxyl groups. These activated carboxyl groups will then be reactive to the primary amine groups on the proteins being coupled. The beads used throughout this study are 4.5 μm carboxylate microspheres (Polysciences Inc PA USA) Cat #17140-5. The polylink protein coupling kit Cat # 24350-1 was used for all the coupling reactions which contains: Coupling buffer, storage buffer and carbodiimide (EDAC).

The procedure for coupling fluorescent proteins to the beads is as follows: Pipette 0.5 ml of the bead suspension into a 1.5 ml

microcentrifuge tube. Centrifuge in a benchtop microcentrifuge at 2000 *g* for 5 min. Carefully aspirate the supernatant then resuspend in 0.5 ml coupling buffer. Repeat the previous wash step. Resuspend the beads in 0.17 ml of polylink coupling buffer. Prepare fresh 10 mg EDAC in 50 µl polylink coupling buffer. Add 20 µl to the bead suspension. Incubate for 10 min at room temperature. After vortexing briefly, add 250 µg fluorescent protein. Incubate at room temperature on a rotary mixer for 4 h. Centrifuge in a benchtop microcentrifuge at 2000 *g* for 5 min. Aspirate supernatant and resuspend in 0.5 ml Polylink storage and wash buffer. Centrifuge in a benchtop microcentrifuge at 2000 *g* for 5 min. Repeat the previous wash step. Carefully aspirate the supernatant then resuspend beads in 1 ml of polylink storage buffer. Add 50 µl of the bead suspension to a 5 ml dropper bottle, add 4 ml dPBS with 0.02% w/v Sodium azide. The bead preparation can be stored at 4°C for 12 months or more. One such reaction can produce twenty 4 ml bottles each of which can yield about 100 tests. Blank beads can be prepared by diluting 25 µl stock beads with 4 ml dPBS with 0.1% Bovine serum albumin and 0.02% w/v Sodium azide.

The same procedure can be used to produce antibody capture beads by coupling polyclonal Goat anti- Rat IgG or Goat anti Mouse IgG (Thermo Scientific #31220 and #31160 respectively). In the case of coupling antibodies 180 µl of antibody (about 360 µg of protein) is added. The diluted antibody capture beads can be stored at 4°C for at least 12 months.

3.1 | Flow cytometry

The FP coupled beads require no preparation but should be thoroughly resuspended before use. All samples were analyzed on a Cytek Aurora™ full spectrum flow cytometer having either four lasers, 488, 561, 640 and 405 nm or five lasers, 488, 561, 640, 405 and 355 nm. One drop of each of the FP coupled beads was dispensed into a 12 × 75 mm tube along with 200 µl PBS. One drop of the blank beads was dispensed into a 12 × 75 mm tube along with 200 µl PBS.

For using the antibody capture beads, one drop of the bead suspension was added to a 12 × 75 mm tube along with 1 µl of conjugated antibody. Samples were incubated in the dark for 10 min. For a universal negative, samples were diluted with 200 µl PBS without washing and used immediately. If the samples were going to have blank beads added to the tube, the labeled beads were washed by adding 1 ml of PBS and centrifuging at 500 *g* for 5 min. The supernatant was removed and the beads resuspended in 200 µl PBS.

For those not familiar with the Aurora platform, the different detectors are labeled by laser (V for 405 nm, B for 488 nm, YG or Y for the 561 nm and R for the 640 nm lasers). The detectors are numbered with the lowest number being the lowest wavelength detector. For example, the V1 detector is the lowest wavelength detector off the 405 nm laser with a bandpass filter allowing light from 420 to 435 nm to pass. Details of the wavelengths of the various detectors can be found on the Cytek Bio website (https://cytekbio.com/pages/user-guides).

At all times the gain setting of the fluorescence detectors were the Cytek Assay Setting with only the gains for the FSC and SSC altered to allow for the differences on size of the particles used.

Spectroflo™ v2.2.0.4 was used for data acquisition. All the FP coupled beads were initially used as reference controls and the blank beads were run as a universal negative. The raw data can be exported to give a spectral signature, unmixed data can be used to evaluate the spillover spread of each FP into the other channels. Spillover spread was evaluated using Flowjo™ software (BD Biosciences, Sa Jose, CA USA).

After unmixing, various combinations of the FP were added together and run as separate samples. Initially all eight FP were run as reference controls but the spillover spreading was unacceptable, so the experiment was duplicated, tdTomato and mCherry were deleted and unmixing was repeated.

4 | COMPARING CELL BOUND FP FLUORESCENCE, BEAD BOUND FP FLUORESCENCE AND FREE PROTEIN FLUORESCENCE

On the Aurora, the median fluorescence intensity of each of the raw channels can be exported out of Spectroflo for each of the FP coupled samples as well as the blank beads. The data can be opened in Excel or some other graphing software such that each column is a different raw channel and each row is a different FP sample. The values from the blank can be subtracted from each FP row to remove the spectral characteristics of the blank beads. The median values can be normalized such that highest value of the entire spectral signature is one. The same procedure can be used for cells expressing FPs. The negative cell signature can be subtracted from the FP expressing cells and the data normalized. This allows the FP expressing beads and cells to be compared on the same scale. A metric referred to as the difference index (DI) is used demonstrate the difference between two signatures. It is calculated as follows: First each signature has the appropriate negative subtracted to remove the influence of the autofluorescence signatures of beads and cells. The signatures are then normalized such that each signature has a maximum value of one. The difference between each of the fluorescence channels is computed and those values are added together. Thus, a value is derived that describes the total difference between two signatures. A correction factor is added to make the difference between two signatures with few similarities equal to one, for this purpose CFP and mCherry were used. Two identical signatures will have a value of zero. The signatures of two batches of mCherry beads coupled 3 years apart can be compared in the same way.

To compare the fluorescent signature of the bead bound FPs and free FP solution, samples were run on a Cary Eclipse fluorescence spectrometer (Agilent Technologies Inc CA USA). One hundred microliters of each of the FP samples was run on the spectrophotometer with the excitation wavelengths chosen to match the lasers on the Aurora instrument. The exported data can be normalized as above to make the highest channel one.

There are many more channels on the spectrophotometer so some manipulation was needed to compare them on a similar scale.

The cells used to compare the bead/cell FP spectra were all HEK 293 T cells transfected with the appropriate FP construct except for the tdTomato which was constitutively expressed in mouse splenocytes.

5 | RESULTS

5.1 | Producing and purifying fluorescent proteins

The FPs were produced by the Monash Protein Production Unit at Monash University, Clayton, Victoria, Australia. Protein was successfully produced from all the plasmids except for iRFP670. This FP precipitated and wasn't usable. All other proteins were delivered at a concentration of 1 mg/ml. The actual concentration of the proteins may have been less than that for some of the proteins as some precipitation was observed. Several attempts were made to dissolve the iRFP670 FP by varying the salt concentration of the buffer but to no avail.

5.2 | Microsphere coupling

All the fluorescent proteins coupled well to the beads as seen in Figure 1. Some of the microspheres are much brighter than others. The eBFP, Cerulean and mKate are less bright than dsRED, mCherry, tdTomato, eGFP and Venus. The differences can be explained by, the brightness of the FP, the sub-optimal excitation of the FPs with the Aurora laser configuration and the differing actual concentration of the purified FPs, as noted above some precipitation was seen in some of the proteins.

As the antibody capture beads are the same beads used for the FPs, no changes to gains are required when running a mixture of fluorescent proteins and antibody coupled reference controls. The same negative can be used for all samples. Typical antibody capture bead histograms are shown in Figure 2.

Carboxyl microsphere coupling is more straightforward than using glutaraldehyde and amine reactive beads and in our hands, carboxyl beads bind more protein (data not shown). Larger microspheres can be used if higher amounts of fluorescence are required. We have experimented with 6, 10 and 20 μm microspheres. The larger microspheres bind more protein but settle more quickly and as there are fewer particles per microgram of bead suspension are thus less economical. The antibody capture beads compare well with commercially available particles. The polyclonal capture antibodies used, bind both light chains equally well but are species specific with little cross reactivity with other species such as hamster.

5.3 | Spectra comparison of FP coupled beads, FP expressing cells and free FP

The spectral signature of data generated by the Cytek Aurora and free fluorescent proteins can be compared by looking at Figure 3. The fluorescent signatures between the two platforms are similar

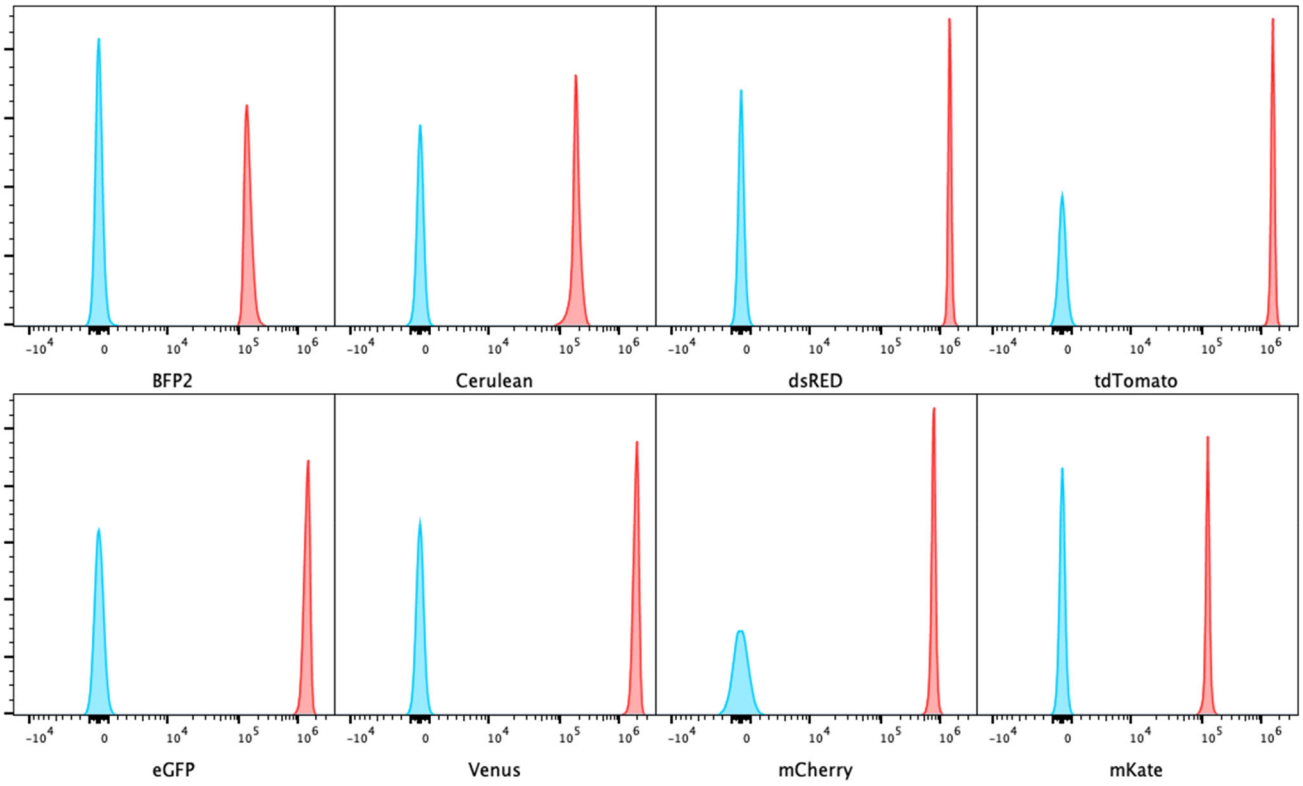


FIGURE 1 Each fluorescent protein coupled to the carboxyl microspheres. The uncoupled beads are shown in blue

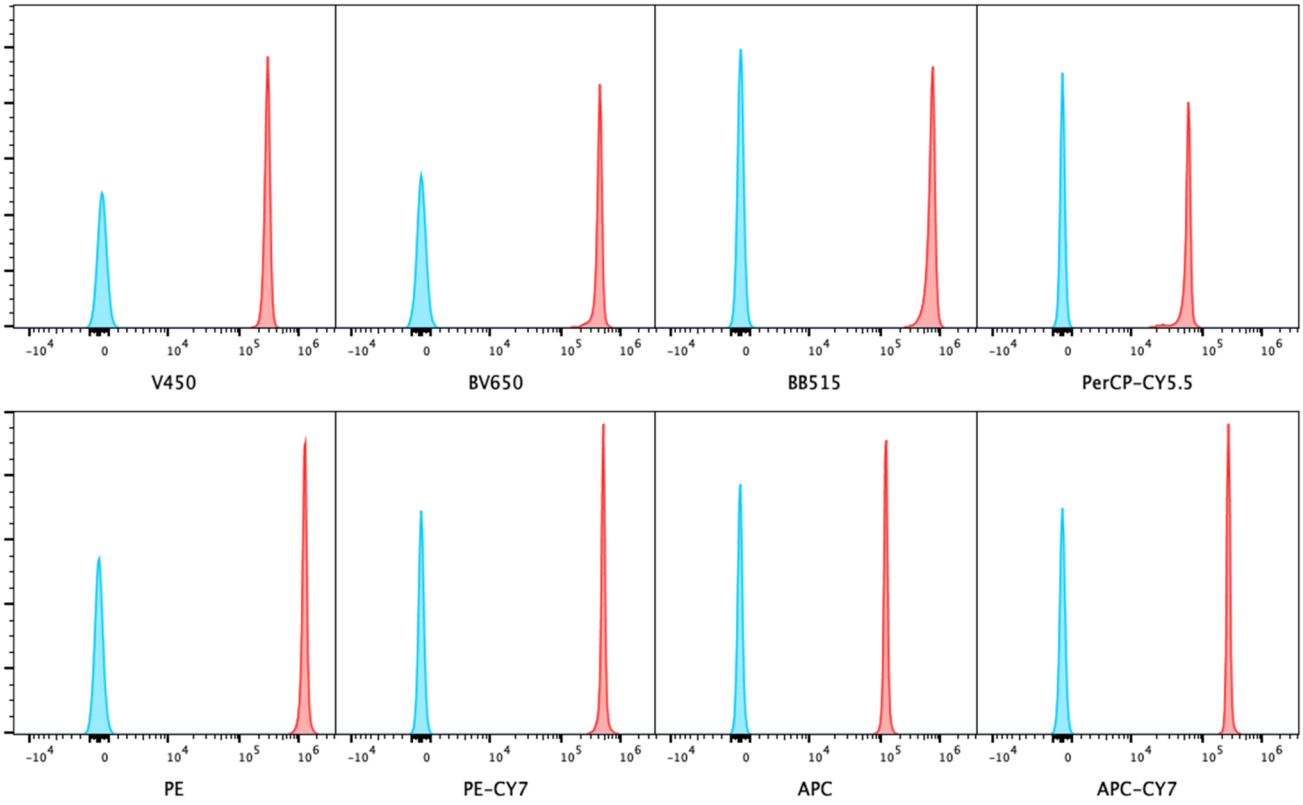


FIGURE 2 Typical antibody coupling with goat-anti rat IgG coupled carboxyl microspheres. The uncoupled beads are shown in blue. All antibodies are Rat anti mouse anti-CD4 (Clone GK1.5) (BD Biosciences, CA USA)

but not identical. The detector arrays on the Aurora do have gaps in the wavelengths detected to avoid scattered light from the lasers. This will distort the spectral signature somewhat and the gains generally used on the Cytek Aurora (Cytek assay setting) are optimized to allow optimal unmixing of multiple antibody fluorochromes, this will explain some of the differences between the Aurora generated data and the spectrophotometer generated data. In particular, mKate despite subtracting the signature of the uncoupled polystyrene particles, does have a small peak in the violet channels of the Aurora which is absent when looking at the signature of the free FP on the spectrophotometer. The signatures of FP coupled beads and FP expressing cells were found to be very similar as seen in Figure 4. The signatures of the uncoupled beads and FP negative cells are subtracted from the FP coupled beads and FP expressing cells respectively. It should be noted that the negative cell signature subtracted from the FP expressing cells should be from identical cells with an identical autofluorescence signature. This is not always possible which explains some of the differences observed. The difference between the beads and cell signatures is shown as the DI and is shown on each plot. The old and the new mCherry coupled beads have a very low DI of 0.02 as they are almost identical. The signatures CFP and mCherry were chosen as signatures with few similarities and the correction factor was chosen to give a DI value of one. Figure 5 shows the results of using FP coupled beads as reference controls. There are some unmixing errors which result in differences in median values in the spillover channels between the unlabelled and labeled particles. The median of both unlabeled and FP expressing cells are shown by the red lines.

5.4 | Using multiple fluorescent proteins together

We demonstrate in Figure 6 that six FPs can be used together with ease, with an acceptable level of spillover spread. If UV excited FPs or far-red excited FPs were available it would be simple to add a few more. The eGFP, Venus, dsRED and tdTomato beads are very bright. If cells are less bright that the beads, the spreading would look less severe. The FP coupled beads can be used in the planning stages of an experiment to ascertain which FPs can be used together along with which antibodies.

5.5 | Trouble shooting

Anecdotally fluorescent proteins are difficult both to purify and couple to beads. We found no such difficulties. Despite being expressed in *E. coli* the spectral signatures were almost identical to those expressed by mammalian cells. Even the dimeric proteins showed the same spectral signatures as expected. Expressing and purifying proteins in *E. coli* is both quicker and less expensive that expressing in insect or mammalian cells. We did experience some slight precipitation of several FPs upon delivery and after several years stored in the

FIGURE 3 Spectral signature for eBFP2, cerulean, eGFP and mVenus from (A) the Aurora and (b) the Cary eclipse fluorescence spectrometer. Spectral signature for dsRED, tdTomato, mCherry and mKate from (C) the Aurora and (D) the Cary eclipse fluorescence spectrometer

fridge in pure dPBS. I would suggest freezing the proteins in 50% glycerol. We have experienced very heterogenous coupling of FPs when too little protein is added. Presumably, the protein is absorbed so rapidly that some beads get little exposure to free FP. This results in a histogram with a "smear" rather than a single homogenous peak. The solution is to add more protein.

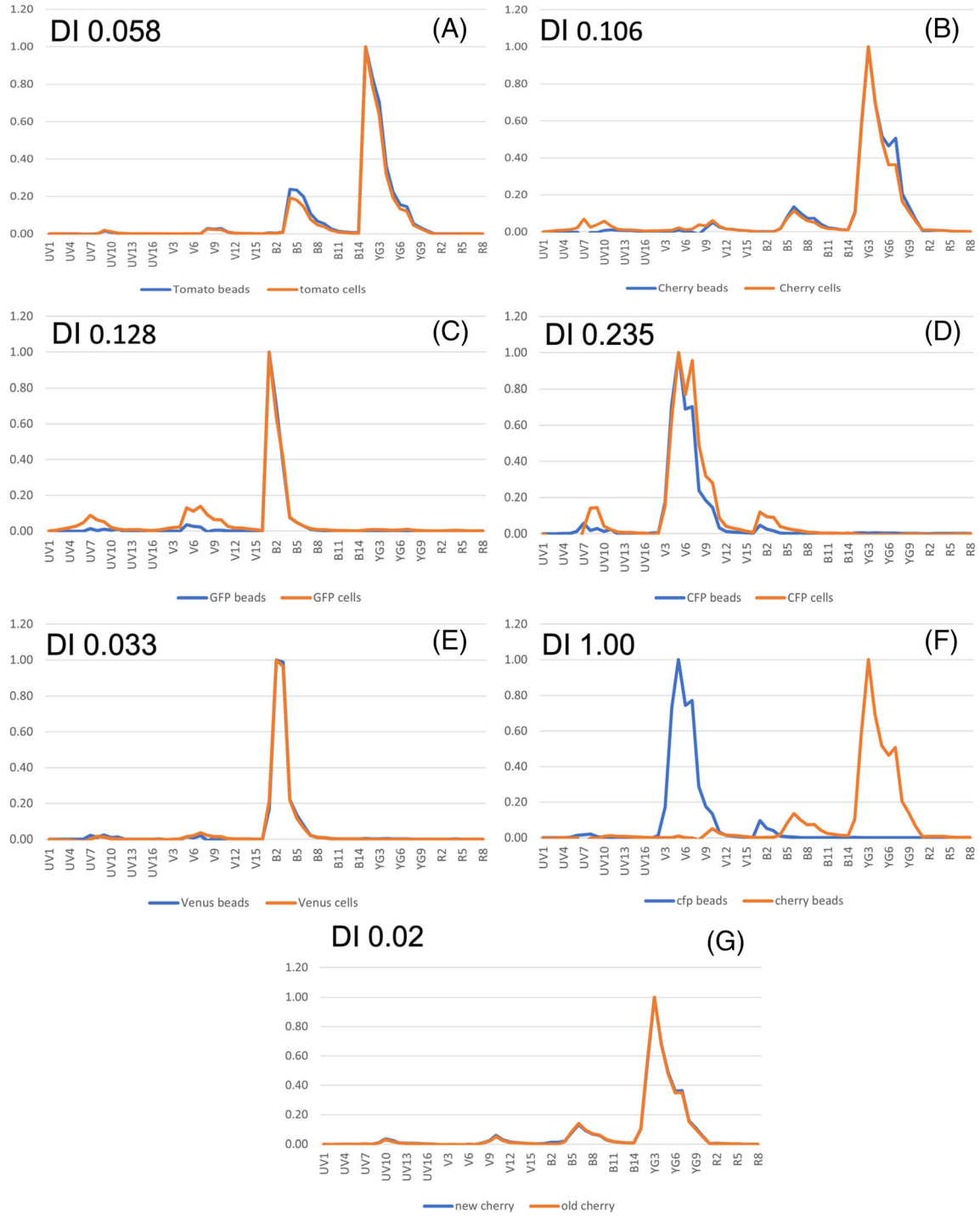


FIGURE 4 Overlay of spectral signature of beads and cells for (A) tdTomato, (B) mCherry, (c) eGFP, (d) CFP, and (e) Venus, (f) CFP and mCherry beads and (g) freshly coupled mCherry beads with mCherry beads coupled 3 years previously. The Difference index (DI) is on the top left of all plots

6 | DISCUSSION

Using antibody capture beads as single color controls has been standard practice in many labs for many years. Carboxylate beads offer simple coupling to proteins, despite the bead binding the amine terminus of immunoglobulin, the coupled antibodies capture fluorescent antibodies well. Using cells for single color controls is preferable but is not always practical, as for instance, the target cell population may be so rare that accumulating a large enough data file is not possible. Fluorescent protein coupled beads are a very convenient option to having FP labeled cells, the particles are ready to use, stable and inexpensive to make. While coupling FPs to beads is a relatively trivial process, the

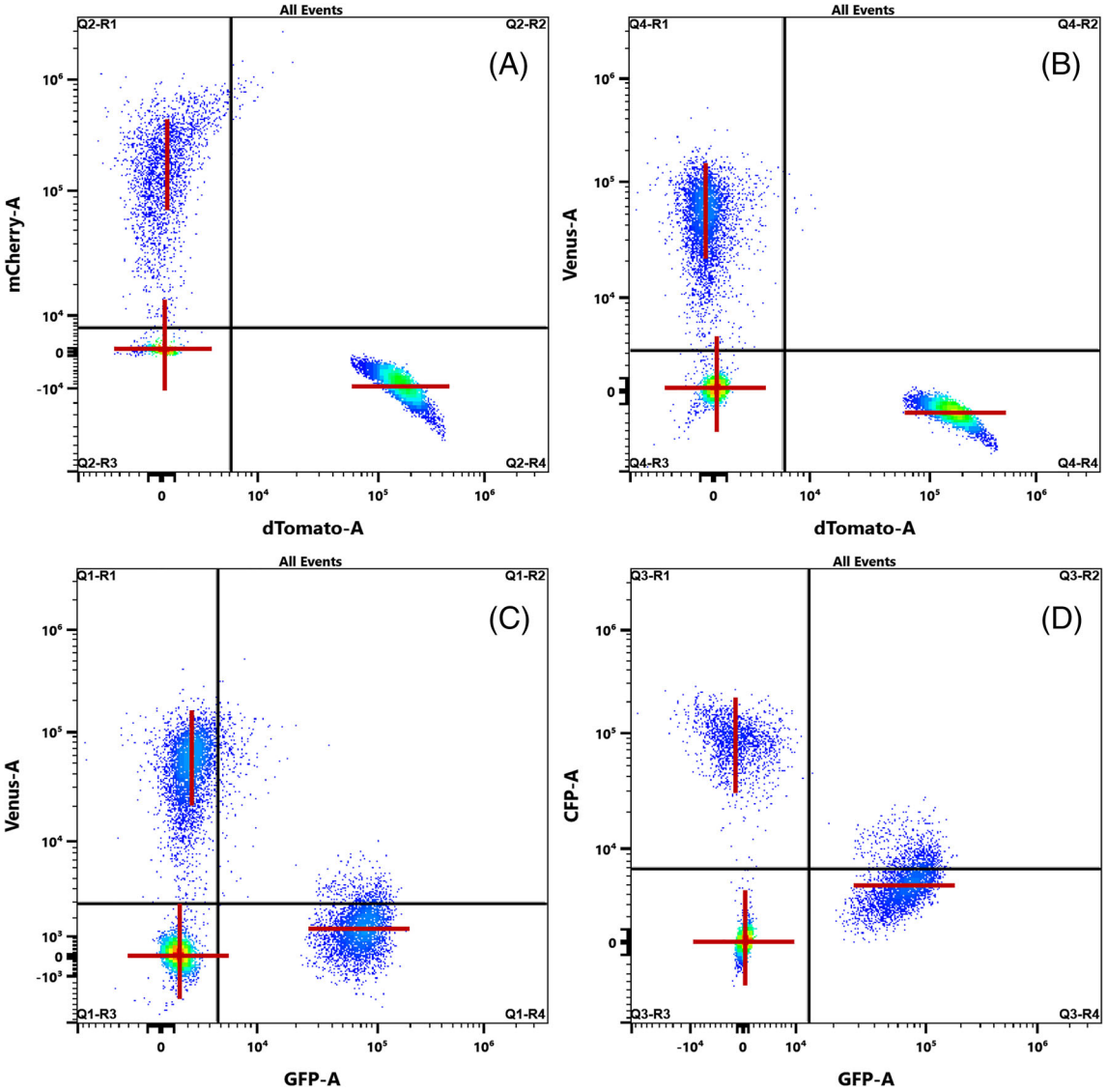


FIGURE 5 FP coupled beads are used as reference controls, the resulting unmixing shows some unmixing errors when looking at combinations of FPs expressed in cells. The medians are shown as red lines for both unlabelled and FP expressing cells. Each dotplot is a concatenated FCS file from FCS files of unlabelled cells and each of the FP expressing cells

difficulty is producing the purified FPs. This may have to be outsourced by using a protein production facility. One milligram of purified FP is enough protein to produce about 10,000 tests. In this study eight fluorescent proteins were successfully produced, purified and coupled to carboxylate polystyrene microspheres. All of the FPs were produced in *E. coli* but the fluorescent characteristics resulting from these proteins appear to be very similar to FPs expressed in eukaryotic cells.

When the FP coupled beads are used as reference controls often some unmixing errors are observed as shown in Figure 5. These unmixing errors may be the result of the coupling process or the FPs coupled to the beads did not always match exactly the FPs expressed the cells studies, for instance tdTomato was coupled to beads but the cells used expressed dTomato and Cerulean was coupled to beads but eCFP was expressed in the cells used. Producing the FPs in *E. coli* is not thought to change the spectral characteristics the proteins There are other methods of coupling proteins to beads which will be investigated to see if the signatures of the FP coupled beads and cells can be make perfect. The Cytek Aurora can detect very small changes to the fluorescence signature, for instance dsRED and tdTomato have very similar signatures but can be separated on the Aurora albeit with a high level of spread.

Mostly monomeric FPs were chosen as it was thought the production would be more straightforward and they were also the most requested FPs at our institute. There is no reason why far-red and UV excited FPs could not be produced in the same way and there are plans to do this in the future. Antibody capture beads have been produced in the Flow Cytometry facility at WEHI for the last 8 years and are sold to investigators for a small fee to cover production. One bottle of 5 ml of 4.5 μm microspheres from the manufacturer can produce about 100 kits of positive and blank beads and each kit can yield about 100 tests.

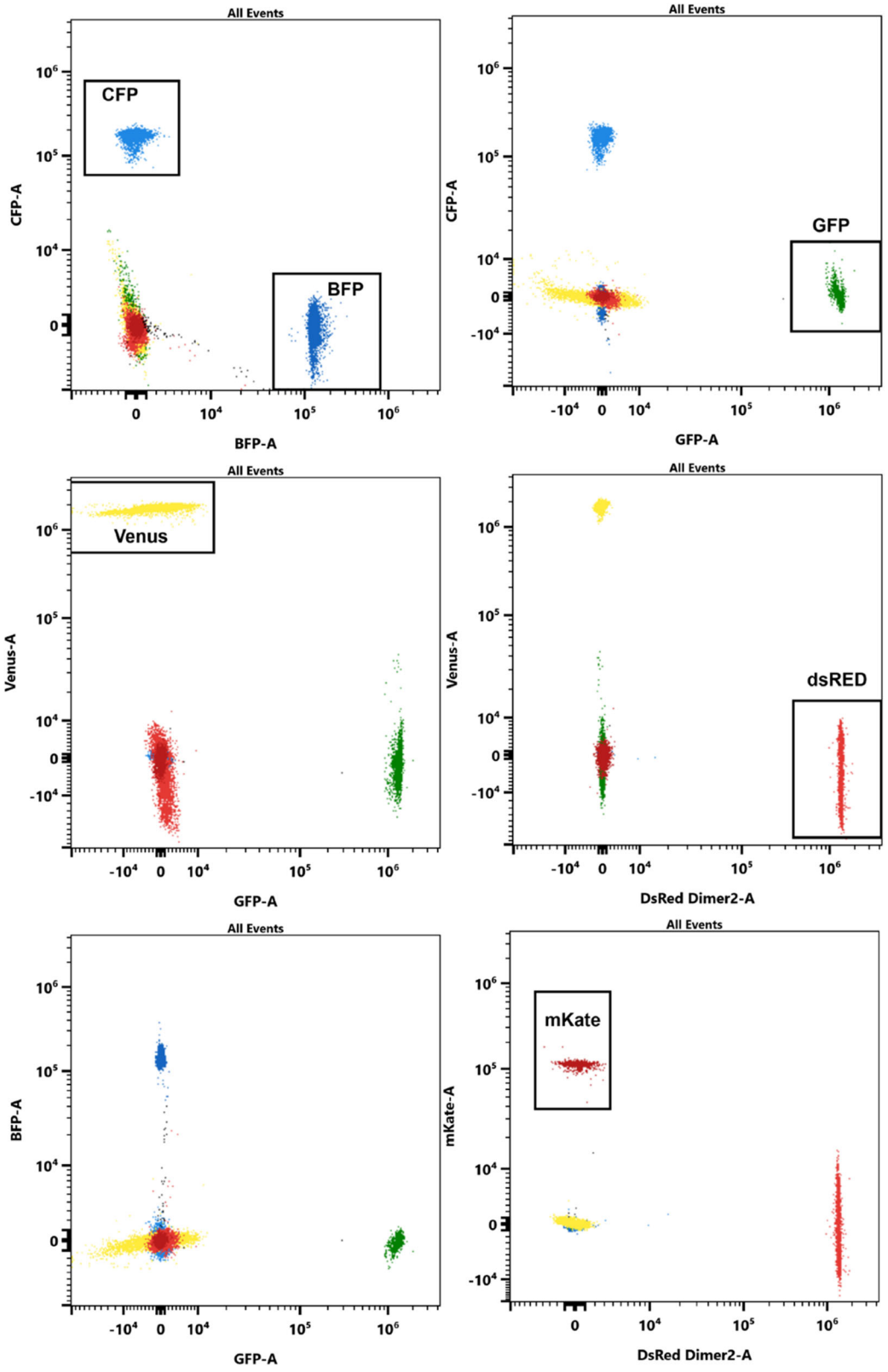


FIGURE 6 Mixture of eGFP, Venus, eBFP2, cerulean, dsRED, and mKate coupled microspheres. Six of the FP coupled beads can be unmixed and separate with an acceptable level of spread on a Cytek Aurora

FP coupled beads and antibody capture beads can be used in the planning stage of an experiment, reference controls from both fluorescent proteins and antibody fluorochromes can be acquired on the aurora, unmixed then the spillover spreading matrix (SSM) can be used to check the spreading. If there are unacceptable levels of spreading, a FP or antibody fluorochrome can be substituted and the samples re-unmixed.

In this study we demonstrate that it is possible to couple fluorescent proteins to carboxyl polystyrene microspheres using a standard protocol. We demonstrate that the spectral signatures of beads and cells are very similar for several FPs. This suggests that the fluorescent signature of the FPs are distorted little by coupling to polystyrene beads making FP coupled beads a convenient surrogate for FP expressing cells. It was found that some red fluorescent proteins, dsRED and tdTomato have signatures so similar that they should not be used together. Likewise, mCherry and mKate are close spectrally resulting in a large spillover spread. We demonstrate that six fluorescent proteins can be used together without difficulty. The brightness of some of the FP coupled beads makes the apparent spreading look worse than it would if the beads were dimmer. With the addition of some UV and far-red excited FPs it should be possible to separate 10 fluorescent proteins. Some unmixing errors were incurred when using FP coupled beads as reference controls so some caution is required when using them, but they represent a convenient alternative to FP expressing cells.

ACKNOWLEDGMENTS

I would like to thank the Monash Protein Production Facility for expressing and purifying all the fluorescent proteins used in this study except eGFP. I would like to thank David Miller and Jacqui Gulbis Structural Biology division, Walter and Eliza Hall Institute for Medical Research (WEHI) for the purified eGFP and Dr. Mona Radwan, Inflammation Division, Walter and Eliza Hall Institute for Medical Research (WEHI) for providing FP transfected HEK 293 T cells.

CONFLICT OF INTEREST

The author has no conflicts of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/cyto.a.24557.

ORCID

Simon Monard https://orcid.org/0000-0003-4447-2256

REFERENCES

1. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. Science. 1994;263:802–5.
2. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the Aequorea victoria green-fluorescent protein. Gene. 1992;111:229–33.
3. Shimomura O, Johnson FH, Saiga Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan. Aequorea J Cell Comp Physiol. 1962;59:223–39.
4. Lambert TJ. FPbase: a community-editable fluorescent protein database. Nat Methods. 2019;16:277–8.
5. Abe T, Sakaue-Sawano A, Kiyonari H, Shioi G, Inoue K, Horiuchi T, et al. Visualization of cell cycle in mouse embryos with Fucci2 reporter directed by Rosa26 promoter. Development. 2013;140:237–46.
6. Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell. 2008;132:487–98.
7. Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA. Fluorescent proteins and their applications in imaging living cells and tissues. Physiol Rev. 2010;90:1103–63.
8. Rodriguez EA, Campbell RE, Lin JY, Lin MZ, Miyawaki A, Palmer AE, et al. The growing and glowing toolbox of fluorescent and photoactive proteins. Trends Biochem Sci. 2017;42:111–29.
9. Lippincott-Schwartz J, Patterson GH. Development and use of fluorescent protein markers in living cells. Science. 2003;300:87–91.
10. Jolley ME, Wang CH, Ekenberg SJ, Zuelke MS, Kelso DM. Particle concentration fluorescence immunoassay (PCFIA): a new, rapid immunoassay technique with high sensitivity. J Immunol Methods. 1984;67:21–35.
11. Novo D, Gregori G, Rajwa B. Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices. Cytometry A. 2013;83:508–20.
12. Staros JV, Wright RW, Swingle DM. Enhancement by N-hydroxysulfosuccinimide of water-soluble carbodiimide-mediated coupling reactions. Anal Biochem. 1986;156:220–2.
13. Bangs Laboratories Inc. Covalent Coupling. 9025 Technology Dr. Fishers, IN 46038-2886. TechNote 205; 2013.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.